High-SNR Capacity of AWGN Channels with Generic Alphabet Constraints

A DISSERTATION PRESENTED
BY
IAN WEINER
TO
THE HARVARD JOHN A. PAULSON SCHOOL OF ENGINEERING AND APPLIED SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
APPLIED MATHEMATICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
JANUARY 2017

This material is based upon work supported under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

Thesis advisor: Professor Vahid Tarokh Ian Weiner

High-SNR Capacity of AWGN Channels with Generic Alphabet Constraints

Abstract

We present a generalized notion of entropy taken with respect to a measure in a coordinate-independent manner and prove several novel entropy convergence theorems. A particular focus is entropy of random variables on smooth submanifolds of $\mathbb{R}^N$.

We apply these results to computing the information capacity of an AWGN channel whose alphabet is constrained to an n-dimensional smooth submanifold of $\mathbb{R}^N$. Such submanifolds are shown to arise naturally when coding alphabets in $\mathbb{R}^N$ are subjected to a set of smooth constraint functions. The asymptotic capacity in the high-SNR limit is computed for such AWGN channels with manifold constraints in two variants: a compact alphabet manifold, and a non-compact scale-invariant alphabet manifold with an additional average power constraint on the input distribution. The high-SNR capacity expression resembles Shannon's famous Gaussian channel capacity formula, with an additional constant term determined by the geometry of the alphabet constraint manifold– namely, a volume derived from the manifold.

Thesis advisor: Professor Vahid Tarokh Ian Weiner

We apply the above theory in a study of the channel capacity of radar pulse waveforms. In our model, each radar pulse also constitutes a code letter for transmission of information. It is desirable in this context to constrain the alphabet of waveforms to those particularly suited to efficient and effective radar signal processing, giving rise to a channel described by the above work. We numerically compute the volume component of our asymptotic capacity expression for a plausible range of performance characteristics of the radar signal processing. We plot curves that show the inherent trade-off for our radar between signal processing performance and channel capacity.

Contents

Acknowledgments

This work could not have been done without the tremendous support, encouragement, and sacrifice of my wife Sara, our daughters Melanie and Libby, my father Jerry, and my mother-in-law Ellen.

The financial and technical support provided by Lincoln Laboratory through the Lincoln Scholars program has been extraordinary. I am particularly indebted to my Lincoln mentors Steve Krich and Christ Richmond, and the enthusiastic support of my group leader Geordi Borsari, in addition to many others.

I owe a great deal to Vahid for his wonderful support and advice. I must also thank Yue Lu, Li Na, and Roger Brockett of SEAS, and Lizhong Zheng and Muriel Médard of MIT for serving on my committees and offering their thoughtful feedback.

Glossary of Notation

Basic Notation

$k \in \{0, 1, \dots\}$, $v \in \mathbb{R}^k$, $S \subset \mathbb{R}^k$, $a \in \mathbb{R}, r \geq 0$.

- $|v|$ is the standard Euclidean vector norm.
- $|S|$ is the standard k-dimensional Euclidean volume of S.
- $B_r^k(v_0) := \{v \in \mathbb{R}^k \colon |v - v_0| < r\}$, the open ball of radius r in $\mathbb{R}^k$ centered at v_0.
  $B_r^k := B_r^k(0)$. $B^k := B_1^k(0)$.
- $B_r^k(S) := \bigcup_{v \in S} B_r^k(v)$, all points within Euclidean distance r of the set S.
- $\omega_k := \left|B_1^k\right| = \pi^{k/2}[\Gamma(1 + k/2)]^{-1}$ ($\omega_0 \equiv 1$).
- $\kappa_{N,k} := \frac{N^N \Gamma(1+k/2)\Gamma(1+(N-k)/2)}{k^k (N-k)^{N-k}\Gamma(1+N/2)} = \frac{N^N \omega_N}{k^k \omega_k (N-k)^{N-k} \omega_{N-k}}$.
- $\partial B_r^{k+1} \equiv S_r^k := \{v \in \mathbb{R}^{k+1} \colon |v| = r\}$, the k-sphere of radius r.
- $aS := \{av \colon v \in S\}$.
- $a_1 \vee a_2 := \max\{a_1, a_2\}$ and $a_1 \wedge a_2 := \min\{a_1, a_2\}$.
- We take log to base e unless otherwise noted; $\log^+ a := (\log a) \vee 0$.

Geometric Notation

$\mathcal{W}$ is a smooth n-dimensional submanifold of $\mathbb{R}^N$, $w \in \mathcal{W}$, τ is a tangent vector at a point.

- $d_{\mathcal{W}}(w_0, w_1)$ is the geodesic distance between w_0 and w_1 ($= \infty$ if there is no connecting geodesic).
- $B_r^{\mathcal{W}}(w_0) := \{w \in \mathcal{W} \colon d_{\mathcal{W}}(w_0, w) < r\}$, the geodesic ball at w_0 of radius r.
- V^n is the n-dimensional volume measure induced on $\mathcal{W}$ by the Euclidean metric of $\mathbb{R}^N$.
- $J_w = \frac{dV^n}{dm^n}$, the Jacobian factor in a geodesic normal coordinate system centered on $w \in \mathcal{W}$.
- $\Theta = \frac{dm^N}{dm^{n'} dV^n}$ is the Jacobian factor for Euclidean N-volume in the tubular parameterization.

Notation for Measures, Functions, and Norms

$(\mathcal{M}, \Sigma, \mu)$ is a σ-finite measure space with $\mu \geq 0$, $S \in \Sigma$ is a measurable set, P a probability measure on $\mathcal{M}$, $f \colon \mathcal{M} \to [-\infty, \infty]$ is μ-measurable, $b \in [1, \infty]$.

- $\mathcal{P}(\mathcal{M})$ is the set of positive measures on $\mathcal{M}$.
- $\hat{\mathcal{P}}(\mathcal{M})$ is the set of probability measures on $\mathcal{M}$.
- $\mathcal{P}(\mu) := \{\mu\text{-measurable } f \colon f \geq 0\}$.
- $\hat{\mathcal{P}}(\mu) := \{f \in \mathcal{P}(\mu) \colon \|f\|_1 = 1\}$, probability densities w.r.t. μ.

- [Abuse of notation] $P \in \hat{\mathcal{P}}(\mu)$ means that $P \ll \mu$ and $\frac{dP}{d\mu} \in \hat{\mathcal{P}}(\mu)$.
- $P \perp Q$ means the probability measures P and Q are independent.
- 1_S is the indicator function of the set S.
- $\|f\|_b$ is the $L^b(\mu)$ norm of f.
- $\|f\|_{b;S} := \|1_S f\|_b$.
- $L^1_+(\mu) := \{f \in L^1(\mu) \colon f \geq 0\}$.

Notation Related to the Normal Distribution

- $\varphi_{k,\varepsilon}(r) := (2\pi\varepsilon^2)^{-k/2} e^{-r^2/2\varepsilon^2}$, the Gaussian pdf on $\mathbb{R}^k$, with zero mean and variance $\varepsilon^2 I_k$, evaluated at $|v| = r$.
- $\chi_{k,\varepsilon}(r) := k\omega_k r^{k-1} \varphi^k_\varepsilon(r)$, the pdf of $|Z|$ when $Z \sim \mathcal{N}(0, \varepsilon^2 I_k)$.
- $\Phi_{n,R} := \int_0^R \chi_{k,\varepsilon}$, the probability of $|Z| \leq R$ when $Z \sim \mathcal{N}(0, \varepsilon^2 I_k)$.
- $\varphi^{(R)}_{k,\varepsilon}(r) := \Phi^{-1}_{n,R/\varepsilon} 1_{[0,R]}(r) \varphi_{n,\varepsilon}(r)$
- $\chi^{(R)}_{k,\varepsilon} := \Phi^{-1}_{n,R/\varepsilon} 1_{[0,R]} \chi_{k,\varepsilon}$, the $\chi_{k,\varepsilon}$ pdf conditioned on $r \leq R$.

1

Introduction

This dissertation consists of three contributions to the Information Theory literature, each leveraging the results of its predecessor.

1.1 Generalized Entropy

The first contribution is the investigation of a generalized definition of entropy with respect to a mathematical measure, which subsumes both the discrete and differential entropy of classical information theory. Specifically, for a probability measure P_X and a positive measure μ, if the Radon-Nikodym derivative $\frac{dP}{d\mu}$ exists, we define

$$h_\mu(P_X) := -\mathbb{E}\left[\log \frac{dP}{d\mu}(X)\right]$$

If $d\mu = dx$, the standard Lebesgue measure on $\mathbb{R}$, this is the standard differential entropy, while if μ is the counting measure on $\mathbb{N}$ it reduces to the discrete entropy.

We rigorously prove several powerful theorems in this context. Our focus is the bounding and estimation of entropy differences $|h(p) - h(q)|$, in terms of the difference of probability density functions $|p - q|$, namely its L^b norm for $1 < b \leq \infty$ and L^α semi-norm for $\alpha \in (0, 1)$. We further prove that the L^α semi-norm may be replaced in most cases by a certain weighted norm of the form

$$\|P_X\|_{(\delta);a} := \mathbb{E}\left[(1 + a|X|)^\delta\right]$$

where a, δ are positive real numbers. This particular norm proves to be extremely convenient for our bounds by virtue of its connection to average power constraints. Most of our results appear to be novel to the Information Theory literature, even when reduced to the special cases of the classical discrete and differential entropies.

The primary motivation for our more abstract definition of entropy is *coordinate independence*; It has no reference to a fixed coordinate system. This property is essential to the study of entropy when the natural probability space of interest is a *smooth manifold*, which typically cannot be fully parameterized under any single fixed coordinate system, but rather relies upon a patchwork of local parameterizations. While this may seem esoteric, it is precisely the situation that arises naturally from the channel capacity problem described in the subsequent section, whose solution in the high-

SNR regime with AWGN is the second contribution of this dissertation.

1.2 Channel Capacity with Generic Alphabet Constraints

Before stating the problem of interest, we first recall the standard (real, memoryless) N-dimensional communications channel $\mathbb{R}^N \to \mathbb{R}^N$ with additive Gaussian noise: $Y = X + Z$, with $X, Y, Z \in \mathbb{R}^N$, $Z \sim \mathcal{N}(0, \Sigma)$ with $Z \perp X$ and Σ is the $N \times N$ covariance matrix of $Z \in \mathbb{R}^N$. For example, this channel is often used to model a band-limited communications system of time-bandwidth product $WT \approx N$. In order to avoid confusion it is important here to emphasize that we are taking the perspective of a *fixed* dimension N, with each N-tuple $(X_1, \ldots, X_N)$ collectively representing a single, discretely transmitted letter of a code. A code of length L from this perspective may be considered a vector in $\mathbb{R}^{LN}$. Let us use $X^{(l)} \in \mathbb{R}^N$ to denote the l^{th} letter transmitted, and $X_k^{(l)} \in \mathbb{R}$ to denote its k^{th} component. An *average power constraint* on the transmitted codes is of the form $\sum_{l=1}^{L} \sum_{k=1}^{N} \left| X_k^{(l)} \right|^2 \leq LN\mathsf{P}$, where P is a fixed constant. Writing $|X|^2 \equiv \sum |X_k|^2$, the average power constraint in the limit of $L \to \infty$ is equivalent to the input distribution constraint $\mathbb{E}|X|^2 \leq N\mathsf{P}$. The capacity of this channel with white noise $\Sigma = \varepsilon^2 I_N$ was found by Shannon[12] to be $\frac{N}{2} \log\left(1 + \frac{\mathsf{P}}{\varepsilon^2}\right)$ nats per transmission. Hence, for a sufficiently large code length L, there are codes, with arbitrarily small probability of decoding error, that transmit information at any rate below this, but no higher.

With this starting point, we prove how to rigorously approximate channel capacity in the AWGN case when it is subjected to additional *generic alphabet constraints*. By this we mean that the alphabet of possible X is restricted to a proper subset $\mathcal{X} \subsetneq \mathbb{R}^N$ which is defined by a generic set of *constraint functions* $\mathsf{F}_j \colon \mathbb{R}^N \to \mathbb{R}$ for $j = 1, \ldots, J$

and corresponding *constraint values* b_j. The corresponding alphabet constraints imposed can be any of the form

$$\mathsf{F}_j\left(\frac{X}{|X|}\right) \gtreqless \mathsf{b}_j, \qquad j = 1, 2, \ldots, J \tag{1.2.1}$$

where we are free to choose from the relations $>, \geq, =, \leq$, or $\leq$ individually for each j as appropriate for the application. With these fixed, the alphabet set $\mathcal{X}$ is defined to be all $X \in \mathbb{R}^N$ satisfying (1.2.1). Note that the constraints are scale-independent, so may properly be considered as generic functions on the unit sphere $S^{N-1} \subset \mathbb{R}^N$. The only assumption required on the F_j is that they are smooth (in fact, our results only require them to be C^2, but we assume C^∞ smoothness for simplicity). The choice of the b_j is also permitted to be nearly arbitrary in our formulation, with the understanding that certain choices result in $\mathcal{X} = \emptyset$.

After accounting for some minor technical details, it is shown that the generic form of $\mathcal{X}$ defined by such constraints is an n-dimensional *submanifold* (possibly *with boundary*) of the ambient space $\mathbb{R}^N$, where $0 \leq n \leq N$. (These terms, and much more, are reviewed in Chapter 2) below.) Due to the scale-invariance of the constraints, $\mathcal{X}$ in fact consists of all scalar multiples of an $(n-1)$-dimensional submanifold $\Omega \subset S^{N-1}$.

By bringing techniques and results of differential geometry to bear on our analysis, and considering the entropy $h_{V^n}(P_X)$, defined with respect to V^n, the n-dimensional volume measure of $\mathcal{X}$, we derive the asymptotic capacity of this alphabet-constrained channel in the high-SNR limit, subject to the average power constraint $\mathbb{E}|X|^2 \leq n\mathsf{P}$. For $\varepsilon \ll \mathsf{P}$, we prove that

$$\mathsf{Cap}(\varepsilon) \approx \frac{n}{2}\log\left(1 + \frac{\mathsf{P}}{\varepsilon^2}\right) + \log\frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})}$$

Moreover, a fixed, simple, explicitly defined input distribution P_X is shown to asymptotically achieve this capacity as $\varepsilon \to 0$. Namely, P_X is independent of $\hat{X} := X/|X|$, and $|X|$ is distributed as a χ distribution in n variables with $\mathbb{E}|X|^2 = n\mathsf{P}$.

In the process of proving this, two other notable results are obtained: First, a corresponding asymptotic capacity result for the AWGN channel with arbitrary *compact* alphabet constraint manifold $\mathcal{X}$ (again, possibly with boundary). This result makes no assumptions of scale-invariance, and also does not consider any average power constraint. Second, a general expression for the high-SNR approximation of $h(P_Y)$ in terms of $h_{V^n}(P_X)$ whenever P_X satisfies certain technical "niceness" conditions (which include being twice differentiable, for example).

1.3 Application: High-SNR Capacity of a Radar Waveform Channel

We demonstrate the power of our theoretical results above by applying them to the original question that motivated the work: *How much information can be transmitted in a radar waveform?* This question is motivated by a vision of efficient spectrum sharing between radar systems and wireless communications systems.

We focus our analysis on radars that operate by transmitting a series of discrete, high power pulses, each constituting a code letter in our existing framework, and the dimension N determined by the time-bandwidth product of the pulses. In order to transmit more information, a large alphabet of potential pulses is desirable. On the other hand, the radar signal processing is most effective for a small class of waveforms that possess optimal characteristics for filtering and target detection. Our goal is to quantify this trade-off between these dual missions of radar performance and information transfer.

We constrain our radar waveform alphabet as follows: for a given candidate waveform, radar target range is processed by a linear time-invariant filter; The constraint function on our alphabet quantifies the optimal performance of such filters in terms of gain on target and the reduction of interference due to filter bank cross-correlation (*range sidelobes*). The form of this constraint function is quite difficult to analyze in closed form for the purposes of applying our geometric theory, but straightforward to evaluate numerically for specific radar parameters. Choosing a representative set of parameters, we implement a numerical Monte-Carlo routine to compute the constant term $\log \frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})}$ in our asymptotic capacity expression. From this we obtain a series of plots quantifying the radar performance/information capacity trade-off.

1.4 Structure of the Dissertation

Chapter 2 is a review of smooth manifolds and other terminology and results from differential geometry which will be needed for our work. Most results are standard and offered with references instead of proofs. A few results are proven directly because good references seemed elusive, but we make no claim of originality to those results.

Chapter 3 begins our original work, introducing entropy with respect to a measure. Sections 3.1 and 3.2 are applicable in very general settings. Section 3.4 converts this into results suitable for submanifolds and introduces our notion of a uniform submanifold. This section ends with the "cutoff theorem", which effectively bounds how much of an entropy estimation error may be incurred by restricting our analysis to a convenient *tubular neighborhood* of the input manifold.

Chapter 4 uses the entropy estimation theorems of Section 3.4 to obtain the asymptotic capacity results. A significant portion of this chapter is tedious technical bounds, which we have quarantined in Subsections 4.2.1 and 4.2.2 to avoid cluttering the main results.

Chapter 5 explores the application to the radar waveform channel, beginning with an overview of radar and radar signal processing in Section 5.1. Our numerical methodology and results are presented in Section 5.3.

2

Review of Differential Geometry

In this chapter we briefly cover the terminology and results from differential geometry pertinent to later chapters. Many of the results can be proven in greater generality, but our treatment is specialized to smooth submanifolds of $\mathbb{R}^N$ for simplicity and concreteness, and mappings to and from the manifold are also assumed smooth. Results stated without an explicit reference are standard in many textbooks, for example [3, 9]. A good modern text covering the classical differential geometry of curves and surfaces is [2].

Definition 2.1.1 (Charts and Manifolds)**.**

(i) For $0 \leq m \leq n$, set $\mathbb{H}^n_m := \{x \in \mathbb{R}^n \colon x_k \geq 0 \text{ for } 1 \leq k \leq m\}$.

(ii) Let $w \in \mathcal{W} \subset \mathbb{R}^N$, and suppose $\phi \colon U \to V$ is a smooth mapping between open sets $U, V \in \mathbb{R}^N$, mapping $0 \in U$ to $w \in V$. Furthermore, assume ϕ is invertible with smooth inverse. Fix $0 \leq n \leq N$.

  (a) If $\phi(U \cap \mathbb{R}^n) = V \cap \mathcal{W}$, so that an n-dimensional piece of U maps exactly onto a corresponding piece of $\mathcal{W}$ within V, ϕ is called a *local coordinate system for* $\mathcal{W}$ *centered at* w. This also can be referred to as a *local parameterization of* $\mathcal{W}$ *at* w, or a *local coordinate chart*.

  (b) If there is a $0 \leq m \leq n$ such that $\phi(U \cap \mathbb{H}^n_m) = V \cap \mathcal{W}$, we will call ϕ a *generalized local coordinate system/chart*. If $m \geq 1$ we also call it a *local boundary coordinate system* and call w a *boundary point*.

(iii) If there is a local coordinate system centered at every $w \in \mathcal{W} \subset \mathbb{R}^N$, all with the same dimension n, then $\mathcal{W}$ is a *smooth* n*-dimensional submanifold of* $\mathbb{R}^N$.

(iv) If there is a generalized local coordinate system centered at every $w \in \mathcal{W} \subset \mathbb{R}^N$, all with the same dimension n, then $\mathcal{W}$ is a *smooth* n*-dimensional submanifold of* $\mathbb{R}^N$ *with generalized boundary*. (The boundary charts *need not* all have the same m.)

(v) If $\mathcal{W} \subset \mathbb{R}^N$ is an n-dimensional manifold, the number $n' = N - n$ is called the *codimension*.

Example 2.1.1.

(i) The sphere S^2 is an example of a submanifold of $\mathbb{R}^3$ which cannot be parameterized with a single coordinate chart. Note that the common "spherical coordinate" parameterization fails to be invertible at the poles.

(ii) The closed ball $\bar{B}^2 = \{|x| \leq 1\}$ is an example of a manifold-with-boundary. So is the closed hemisphere $\bar{B}^2 \cap \mathbb{H}^3_1$.

The following classical results let us prove that equality constraint functions generically give rise to manifolds, and a mixture of equality and inequality constraints give rise to manifolds with boundary.

Definition 2.1.2. Let F be a smooth mapping on $\mathbb{R}^N$ taking values in $\mathbb{R}^{n'}$. At each $x \in \mathbb{R}^N$, denote the $n' \times N$ matrix of partial derivatives by DF_x. If $\operatorname{rank}(\mathsf{DF}_x) < n'$, x is called a *critical point* of F, and $\mathsf{b} = \mathsf{F}(x) \in \mathbb{R}^{n'}$ is called a *critical value*. If $\mathsf{b} \in \mathbb{R}^{n'}$ is not a critical value, it is called a *regular value*.

In the next two theorems we assume $N \geq n'$ and put $n := N - n'$.

Theorem 2.1.1 (Sard's Theorem)**.** *If $\mathsf{F}\colon \mathbb{R}^N \to \mathbb{R}^{n'}$ is C^{n+1} then the set of critical values of F has Lebesgue measure zero in $\mathbb{R}^{n'}$.*

This well-known result is proven in [10].

Theorem 2.1.2 (Regular Surfaces)**.** *Let $\mathsf{b} \in \mathbb{R}^{n'}$ be a regular value of $\mathsf{F}\colon \mathbb{R}^N \to \mathbb{R}^{n'}$. Then, $\mathcal{W} := \{x \in \mathbb{R}^N \colon \mathsf{F}(x) = \mathsf{b}\}$ is either the empty set or a smooth submanifold of $\mathbb{R}^N$ of dimension $n := N - n'$.*

This result is standard. See, for example, [3, 9].

We will now show that, if $\mathcal{W} \subset \mathbb{R}^N$ is defined by a set of smooth [in]equality constraints, it is a smooth submanifold (possibly with boundary) of $\mathbb{R}^N$ in the *generic case*:

Theorem 2.1.3. *Let* $\mathsf{F}\colon \mathbb{R}^N \to \mathbb{R}^J$ *be a smooth map whose* j^{th} *component is* F_j*, and* $\mathsf{b} := (\mathsf{b}_1, \ldots, \mathsf{b}_J) \in \mathbb{R}^J$*. For each choice of* b*, define* $\mathcal{W}_\mathsf{b} := \{w \in \mathbb{R}^N : \mathsf{F}(X) \gtreqless \mathsf{b}\}$*. Let* $0 \leq n' \leq J$ *be the number of strict equality constraints imposed, and set* $n := N - n'$*. For each* b *we have three possibilities:*

(a) $\mathcal{W}_\mathsf{b} = \emptyset$*, occurring when the imposed constraints are impossible to satisfy.*

(b) $\mathcal{W}_\mathsf{b}$ *is a smooth submanifold of dimension* n *(possibly with boundary when inequality constraints are used), or*

(c) Neither of the above.

The set of b *for which (c) holds has Lebesgue measure zero in* $\mathbb{R}^J$*.*

Remark. In the special case of linear equality constraints, (a) corresponds to an inhomogeneous system of equations $\mathsf{A}x = \mathsf{b}$ with $\operatorname{rank}(\mathsf{A}) < n'$ and $\mathsf{b} \notin \operatorname{Span}\{\mathsf{A}\}$, hence no solution. (b) corresponds to a system of equations when A has rank n', and (c) corresponds to a system with $\operatorname{rank}(\mathsf{A}) < n'$ and $\mathsf{b} \in \operatorname{Span}\{\mathsf{A}\}$, thus allowing solutions. Note that when $\operatorname{rank}(\mathsf{A}) < n'$, $\operatorname{Span}\{\mathsf{A}\}$ is a proper linear subspace of $\mathbb{R}^{n'}$, hence has Lebesgue measure zero, as required.

Proof. In the case of only equality constraints, the theorem follows immediately from the previous two theorems. To extend this to inequality constraints, break up $\mathcal{W}_\mathsf{b}$ into the disjoint sets under which the $2^{J-n'}$ possible combinations of inequality constraints are active. Each active constraint region corresponds to its own set of equality constraints, hence is itself a smooth submanifold for all but a measure-zero set of b. The union of these exceptional sets is still measure-zero. $\square$

Remark. The moral of this theorem is that a *generic* set of realizable constraints will *almost surely* define a manifold (possibly with boundary). However, there is the possibility of this failing for an exceptional choice for b. One intuitive way to understand the significance of this minor technical caveat is the following: If a specified choice of the constraint values $\mathsf{b} \in \mathbb{R}^J$ happens to *not* give rise to a smooth submanifold, then at least we are *assured* that there are uncountably many alternate choices b' which do give rise to a smooth submanifold. Furthermore, these choices are guaranteed to exist arbitrarily closely to our original $\mathsf{b} \in \mathbb{R}^J$.

2.2 Tangent Vectors, Covariant Derivatives, Geodesics, Exponential Map

Definition 2.2.1 (Tangent/Normal Spaces)**.**

(i) Each point $w \in \mathcal{W}$ has a *tangent space to* $\mathcal{W}$ *at* w, denoted $T_w\mathcal{W}$, which is an n-dimensional real vector space centered at w.

(ii) If $\gamma(t)\colon (-1,1) \to \mathcal{W}$ is a curve on $\mathcal{W}$ with $\gamma(0) = w$, then $\gamma'(0) \in T_w\mathcal{W}$, and $T_w\mathcal{W}$ is the space of all such tangent vectors.

(iii) The orthogonal compliment to $T_w\mathcal{W}$, consisting of vectors based at w that are perpendicular to $T_w\mathcal{W}$ (under the standard inner product on $\mathbb{R}^N$), is the *normal space to* $\mathcal{W}$ *at* w, denoted $N_w\mathcal{W}$ or $T_w^\perp\mathcal{W}$.

(iv) A smoothly varying inner product on the tangent spaces of $\mathcal{W}$ is called a *Riemannian metric*, usually denoted g. The standard inner product of $\mathbb{R}^N$, restricted to the tangent spaces, is the Riemannian metric induced by embedding $\mathcal{W}$ in the ambient space $\mathbb{R}^N$.

(v) A *tangent vector field* is a mapping assigning (smoothly-varying) tangent vectors to some subset of $\mathcal{W}$. If $\gamma\colon (-1,1) \to \mathcal{W}$ is a curve, and $V(t)$ is a smooth mapping from $t \in (-1,1)$ to $V(t) \in T_{\gamma(t)}\mathcal{W}$, we call $V(t)$ *a tangent vector field on* γ.

(vi) If a curve $\gamma(t)$ satisfies $|\gamma'(t)| = 1$ for all t, it is *arc-length parameterized.*

Since $\mathcal{W} \subset \mathbb{R}^N$, tangent vector fields $V(t)$ along a curve on $\mathcal{X}$ can be expressed in two ways: *intrinsically*, in terms of n coordinate directions parameterizing a chart on a neighborhood of $\mathcal{W}$, or *extrinsically*, in terms of the N basis directions of the ambient space.

Definition 2.2.2 (Covariant Derivatives)**.**

(i) Let $V(t)$ be a tangent vector field along the curve $\gamma(t)$. V can be extrinsically represented as $\left(V^1(t), \ldots, V^N(t)\right) \in \mathbb{R}^N$. Let $\bar{D}V(t) := \frac{dV}{dt}(t) \in \mathbb{R}^N$, the component-wise derivative. Even though $V(t) \in T_{\gamma(t)}\mathcal{W}$, in general $\bar{D}V(t) \in T_{\gamma(t)}\mathcal{W} \oplus T^{\perp}_{\gamma(t)}\mathcal{W} \approx \mathbb{R}^N$, i.e. its perpendicular component need not be zero.

(ii) The *covariant derivative of a vector field V along the curve γ* is $DV(t) := \left(\bar{D}V(t)\right)^{\top}$, is the orthogonal projection of $\bar{D}V(t)$ onto $T_{\gamma(t)}\mathcal{W}$, at every t along its definition. (Specific notation can vary; Some authors will write $V'(t)$ instead of $DV(t)$.)

(iii) Instead of being defined only on a curve, suppose U, V are two vector fields defined on an open set $\mathcal{U} \subseteq \mathcal{W}$. At each $w \in \mathcal{U}$, the extrinsic directional derivative of V can be taken in the direction specified by $U(w)$. This is denoted by $\bar{D}_U V$ (or $\bar{\nabla}_U V$). The *covariant derivative of V with respect to U* is again the projection onto the tangent space at each point: $D_U V \equiv \nabla_U V := \left(\bar{\nabla}_U V\right)^\top$.

(iv) Let $\gamma(t)$ be a curve on $\mathcal{W}$. Then $V(t) := \gamma'(t)$ defines a tangent vector field along γ. If $DV(t) = 0$ for all t, γ is called a *geodesic* curve on $\mathcal{W}$. If $w = \gamma(t)$ for some t we say that γ *is a geodesic at w with velocity* $\gamma'(t)$.

Remark.

(i) The covariant derivative is frequently defined differently, through tensor notation that we are trying to avoid delving into here. See, for example, [9, Ch. 4] for a proof that our definition is equivalent.

(ii) From our definition, a geodesic is simply a curve on $\mathcal{W}$ whose extrinsic "acceleration" $\bar{D}\gamma'(t) \equiv \frac{d^2\gamma}{dt^2}(t) \in \mathbb{R}^N$ is always perpendicular to the tangent spaces of $\mathcal{W}$.

For every point $w \in \mathcal{W}$ and unit-length tangent direction $u \in T_w\mathcal{W}$ there is a unique arc-length parameterized geodesic going through w with velocity u. In fact, we can say much more:

Theorem 2.2.1 (Exponential Map)**.** *For every $w \in \mathcal{W}$ there is an open neighborhood $\mathcal{U} \subset \mathcal{W}$ containing w and an open neighborhood $\tilde{\mathcal{U}} \subset T_w\mathcal{W} \approx \mathbb{R}^n$. These open neighborhoods may be chosen such that a bijection $\mathsf{expm}_w : \tilde{\mathcal{U}} \to \mathcal{U}$ between tangent vectors and geodesics through w, can be defined in the following way: $\mathsf{expm}_w(0) = w$, and for any unit-norm $u \in T_w\mathcal{W}$, for $|t|$ small enough to guarantee that $tu \in \tilde{\mathcal{U}}$, we have*

$\mathsf{expm}_w(tu) = \gamma_u(t)$, *where* γ_u *is the unique arc-length parameterized geodesic through* w *with velocity* u *at* w*. Furthermore,* expm_w *and its inverse* expm_w^{-1} *are smooth.*

The map expm_w is called the *the exponential map* due to its historical application as the matrix exponential on manifolds of matrix groups. Note that expm_w maps rays emanating from the origin of $T_w\mathcal{W}$ onto the geodesic curves through w whose velocity at w is determined by the direction of the tangent space ray. The key property of geodesics is that they are length-minimizing:

Theorem 2.2.2 (Geodesics are locally minimal)**.** *If* $\mathcal{U}$ *is a normal neighborhood about* $w \in \mathcal{W}$*, and* $w_2 \in \mathcal{U}$*, then (up to reparameterizations), the geodesic ray connecting* w *to* w_1 *minimizes arc-length among all piecewise differentiable curves connecting those points.*

This is proven in [3].

Definition 2.2.3 (Geodesic Balls and Coordinates)**.**

(i) The *geodesic distance* $d_{\mathcal{W}}(w_1, w_2)$ between $w_1, w_2 \in \mathcal{W}$ is the infimum of the arc-lengths of all geodesic curves starting at w_1 and ending at w_2 (or vice versa, by reversing the curves). If no such connecting curves exist, set $d_{\mathcal{W}}(w_1, w_2) = \infty$.

(ii) Define the *geodesic ball of radius* $\rho > 0$ *at* $w_0 \in \mathcal{W}$ as $B_\rho^{\mathcal{W}}(w_0) := \{w \in \mathcal{W} \colon d_{\mathcal{W}}(w_0, w) < \rho\}$. Note that if ρ is small enough that $B_\rho^n(0) \subset \tilde{\mathcal{U}} \subset T_w\mathcal{W}$, then $B_\rho^{\mathcal{W}}(w) \equiv \mathsf{expm}_w(B_\rho^n(0))$.

(iii) Let $\mathcal{U}$ be a normal neighborhood of $w_0 \in \mathcal{W}$. Choose an orthonormal basis of tangent vectors $\{u_1, \ldots, u_n\}$ for $T_{w_0}\mathcal{W}$, so we can concretely identify $T_{w_0}\mathcal{W}$

with $\mathbb{R}^n$ and define a local parameterization using the maps

$$\check{\tau}(t^1,\ldots,t^n) := \sum_{i=1}^{n} t^i u_i \in T_{w_0}\mathcal{W}, \qquad \check{w}(t^1,\ldots,t^n) := \mathsf{expm}_{w_0}(\check{\tau}(t^1,\ldots,t^n)) \in \mathcal{W}$$

valid on the open set $\check{\mathcal{U}} = \check{w}^{-1}(\mathcal{U}) \subset \mathbb{R}^n$. This is the *geodesic normal coordinate system at w_0 with respect to* $\{u_i\}$ (normal coordinates for short).

(iv) The $(t^1,\ldots,t^n)$ themselves may be parameterized in polar coordinates $\check{t}^i(r,\omega)$ for $r > 0$ and $\omega \in S^{n-1}$. The composition $\big(\check{w} \circ \check{t}\big)(r,\omega)$ will be called *geodesic polar coordinates at w_0*.

2.3 Volume

Definition 2.3.1. Let $\check{w}(t^1,\ldots,t^n)$ be a parameterization mapping $\tilde{\mathcal{U}} \subset \mathbb{R}^n$ onto $\mathcal{U} \subset \mathcal{W}$. Define the components of a real $n \times n$ matrix-valued function $G(t^1,\ldots,t^n)$ on $\tilde{\mathcal{U}}$ by $G_{ij} := \big\langle \frac{\partial \check{w}}{\partial t^i}, \frac{\partial \check{w}}{\partial t^j} \big\rangle$, and a Borel measure V^n on $\mathcal{U}$ by

$$V^n(S) := \int_{\tilde{\mathcal{U}}} (1_S \circ \check{w}) \sqrt{\det G}\, dt^1 \cdots dt^n$$

We use the notation 1_S for the indicator function of the set S, taking the value 1 for points in S and zero otherwise.

Remark. The matrix G represents the Riemannian metric g in the coordinate basis vectors. It is an exercise in linear algebra to show that $\sqrt{\det G}$ is the n-dimensional volume of the parallelepiped spanned by the tangent vectors $\{\frac{\partial \check{w}}{\partial t^i}\}_{i=1}^n$. When $n = 1$, V^n is arc-length, and when $n = N$, $\sqrt{\det G}$ is the Jacobian factor for the Lebesgue measure on $\mathbb{R}^N$.

Theorem 2.3.1. *Each V^n is independent of choice of parameterization $\check{w}$. There is a unique Borel measure defined on all of $\mathcal{W}$ that agrees with the prior definitions on each parameterized subset $\mathcal{U}$.*

Definition 2.3.2. From now on V^n will refer to this unique measure defined on all of $\mathcal{W}$. We will call it *the n-dimensional volume on $\mathcal{W}$ induced by $\mathbb{R}^N$*. When working in a geodesic normal coordinate system about w_0, we will sometimes use the shorthand J_{w_0} in place of $\sqrt{\det G}$.

2.4 Second Fundamental Form, Shape Operator, Curvature

Previously we saw that if U, V are tangent vector fields defined on an open $\mathcal{U} \subseteq \mathcal{W}$, we have the extrinsic directional derivative $\bar{D}_U V$ and covariant derivative $D_U V = \left(\bar{D}_U V\right)^\top$. The normal piece of $\bar{D}_U V$ encodes important information, too:

Definition 2.4.1 (Second Fundamental Form and Shape Operator)**.**

(i) Let $w \in \mathcal{W}$ and U, V be tangent vector fields defined in an open neighborhood containing w. The *second fundamental form at w* is a map $T_w\mathcal{W} \times T_w\mathcal{W} \to T_w^\perp \mathcal{W}$ given by $\mathrm{II}_w(U, V) := \left(\bar{D}_U V\right)^\perp$.

(ii) Closely related is the *shape operator at w*, a linear map $S_{w,N} \colon T_w\mathcal{W} \to T_w\mathcal{W}$ defined for each $w \in \mathcal{W}$ and normal vector field N defined near w, as follows: For each tangent vector field U defined near w, $S_{w,N}U := -\left(\bar{D}_U N\right)^\top$.

"Form" here refers to the bilinear forms of linear algebra. In older terminology, the metric g, giving the inner product on tangent spaces, was referred to as *the first fundamental form of $\mathcal{W}$*. The second fundamental form is, indeed, a symmetric bilinear form, and carries the same information as the shape operators:

Theorem 2.4.1. *$\mathrm{I\!I}_w(U,V)$ is symmetric in U,V: $\mathrm{I\!I}_w(U,V)=\mathrm{I\!I}_w(V,U)$. If $f\colon \mathcal{W}\to\mathbb{R}$, we define the vector field fU by scaling $U(w)$ by the value $f(w)$ at each w. $\mathrm{I\!I}$ is bilinear over functions, in the sense that*

$$\mathrm{I\!I}_w(f_1U_1+f_2U_2,V)=f_1(w)\,\mathrm{I\!I}_w(U_1,V)+f_2(w)\,\mathrm{I\!I}_w(U_2,V)$$

The values $\mathrm{I\!I}_w(U,V)$ and $S_{w,N}U$, formally defined in terms of vector fields, in fact depend only on the point vectors obtained by evaluation at w: $\tau_1=U(w),\tau_2=V(w),\nu=N(w)$. Hence, when appropriate we write $\mathrm{I\!I}_w(\tau_1,\tau_2)$ and $S_{w,\nu}\tau$, with $\tau,\tau_1,\tau_2\in T_w\mathcal{W}$ and $\nu\in T_w^\perp\mathcal{W}$.

The shape operator is self-adjoint. It is equivalent to the 2nd fundamental form in the following sense: When expressed in any orthonormal basis of $T_w\mathcal{W}$, the matrix of the linear operator $S_{w,\nu}$ agrees with that of the bilinear form $(\tau_1,\tau_2)\mapsto\nu\cdot\mathrm{I\!I}_w(\tau_1,\tau_2)$.

We will need the following useful relationship between $\mathrm{I\!I}$ and geodesic curves:

Lemma 2.4.2. *For any geodesic curve $\gamma(t)$ on $\mathcal{W}$,*

(i) $\gamma''(t)=\mathrm{I\!I}_{\gamma(t)}(\gamma'(t),\gamma'(t))\in T^\perp_{\gamma(t)}\mathcal{W}$.

(ii) If $\nu\in T^\perp_{\gamma(t)}(\mathcal{W})$ then for some c between s and t,

$$\nu\cdot[\gamma(s)-\gamma(t)]=\frac{1}{2}(s-t)^2\nu\cdot\mathrm{I\!I}_{\gamma(c)}(\gamma'(c),\gamma'(c))$$

Proof. (i) is evident from the definitions: $\gamma''(t)=\bar{D}_{\gamma'(t)}\bar{D}_{\gamma'(t)}\gamma(t)=\bar{D}_{\gamma'(t)}\gamma'(t)=(\bar{D}_{\gamma'(t)}\gamma'(t))^\perp=\mathrm{I\!I}_{\gamma(t)}(\gamma'(t),\gamma'(t))$. For (ii), consider a 1st order Taylor series of $f(s):=\nu\cdot[\gamma(s)-\gamma(t)]$, centered at t. Since $f(t)=f'(t)=0$ and $f''(s)=\nu\cdot\gamma''(s)$, the result follows by the mean-value form of the 2nd order remainder term. □

Example 2.4.1. Let $r > 0$ and $\mathcal{W} = S^1_r(0) \subset \mathbb{R}^2$. $n = 1, N = 2$. The tangent space at angle θ is spanned by $U(\theta) := (\cos\theta, \sin\theta)$, and the normal space is spanned by $\nu(\theta) := (-\sin\theta, \cos\theta)$, the inward-pointing normal. Set $w_0 = (0, r)$. The curve $\gamma(t) = (r\sin(t/r), r\cos(t/r))$, which is arc-length parameterized, satisfies $\gamma(0) = w_0$, $\gamma'(0) = (1, 0) = rU(w_0)$, and $\gamma''(0) = (0, -1/r) = (1/r)\nu(w_0)$. $\gamma''(0)$ is already orthogonal to $T_{(0,r)}S^1_r \Leftrightarrow D_{\gamma'}\gamma' = 0 \Leftrightarrow \gamma$ is a geodesic curve. We have $\mathrm{II}_{(0,r)}(\gamma'(0), \gamma'(0)) = (1/r)\nu(w_0)$. Indeed, by symmetry, we can conclude that $\mathrm{II}(U, U) = (1/r)\nu$ for all $w \in S^1_r(0)$.

Example 2.4.2. When $n = 2$ and $N = 3$, $\mathcal{W}$ is a surface in $\mathbb{R}^3$. The normal space is 1-dimensional, so if a unit normal vector $\nu \in T^{\perp}_{w_0}\mathcal{W}$ is chosen, we can consider the symmetric bilinear form $(u, v) \mapsto \langle \mathrm{II}_{w_0}(u, v), \nu\rangle$ on $T_{w_0}\mathcal{W}$, or equivalently, the shape operator $S_{w,\nu}$. From linear algebra it is well-known that the self-adjoint $S_{w,\nu}$ has real eigenvectors u_1, u_2 forming an orthonormal basis of $T_w\mathcal{W}$ with eigenvalues λ_1, λ_2. Thus $S_{w,\nu}u_i = \lambda_i u_i \Leftrightarrow \mathrm{II}_{w_0}(u_i, u_j) = \lambda_i\delta_{ij}$. Geometrically this may be interpreted as follows: in a small normal neighborhood about w_0, define geodesics γ_1, γ_2 through w_0 with velocities u_1, u_2, respectively. γ_i may be approximated to 2nd order at w_0 by a circle in the (u_i, ν) plane, that passes through w_0 tangentially to γ_i, with radius $|1/\lambda_i|$ and centered at $w_0 + \nu/\lambda_i$. (In the limiting case of $\lambda_i = 0$ the circle "of infinite radius" is simply a straight line.)

Remark. The $n = 2$, $N = 3$ surface case is classical and was treated by Euler. The approximating circle to a curve is called the *osculating circle* at the point, and the eigenvalues of the 2nd fundamental form are called the *principal curvatures* at the point, although it must be noted that these are *not* curvatures in the sense it is typically meant in modern differential geometry, which we will define in a moment.

On the other hand, the product of the principal curvatures, (or equivalently, the determinant of the shape operator), is called the *Gaussian curvature* of a surface. $K = \lambda_1\lambda_2 = \det S_{w,\nu}$ *is* a curvature in the modern sense (specifically, the unique sectional curvature of the surface at each point; See the definition below.) Geometrically $K(w)$ describes how the surface is bending near point w; If $K(w) > 0$ it looks like an ellipsoid, if $K(w) < 0$ it looks like a saddle point.

The Gaussian curvature was shown by Gauss to depend only on the *intrinsic* geometry of the surface (that is, the metric on tangent spaces), which is not at all obvious, given that II is defined in terms of normal spaces. Geometrically, this means that the Gaussian curvature of a surface is not changed if a surface is transformed in a way that preserves the distances and angles measured on the surface itself. For example, a flat plane, which has $K \equiv 0$ everywhere, may be rolled into a cylinder without distorting the distances or angles on the surface, so a cylinder also has $K \equiv 0$ (which can also be proven directly.) In contrast is the following canonical "real-world" observation: the sphere S_r^2 has both principal curvatures equal to $1/r$ (relative to an inward-pointing normal; both $-1/r$ rel. an outward normal). Therefore S_r^2 has constant positive Gaussian curvature $1/r^2$. This proves that it is impossible to make a flat map of the Earth (or even any extensive solid angle of the Earth's surface) without introducing distortions in some of the lengths, areas, and angles being represented. Hence, the existence of dozens of map projection methods, most of which portray the planet's land mass as highly concentrated in Antarctica and Greenland.

Definition 2.4.2. Let $\mathcal{W}$ be any smooth n-dimensional submanifold of $\mathbb{R}^N$, $w \in \mathcal{W}$, and let $\sigma_w \subset T_w\mathcal{W}$ be a 2-dimensional subspace of the tangent space at w. Choose an orthonormal basis u_1, u_2 for σ_w. The *sectional curvature of* σ_w is

$$K(\sigma_w) := \langle \mathrm{II}_w(u_1, u_1), \mathrm{II}_w(u_2, u_2) \rangle - \langle \mathrm{II}_w(u_1, u_2), \mathrm{II}_w(u_1, u_2) \rangle$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $T_w^\perp \mathcal{W}$ induced by the dot product on $\mathbb{R}^N$.

Below, it will be helpful to also define the following: II_{σ_w} is II_w restricted to σ_w, and $S_{\sigma_w,\nu} \colon \sigma_w \to \sigma_w$ is $S_{w,\nu}$ restricted to σ_w and having its output restricted to σ_w by orthogonal projection.

Remark. $K(\sigma_w)$ is independent of the choice of orthonormal basis $\{u_1, u_2\}$, and depends only upon intrinsic geometric quantities computed from the Riemannian metric g, despite being defined in terms of the (non-intrinsic) 2nd fundamental form.

When $n = 2$ and $N = 3$, viewing II_w as a real-valued bilinear form as above, our definition of sectional curvature reduces to a 2×2 matrix determinant, giving the product of the eigenvalues, showing that sectional curvature of a surface is its Gaussian curvature. The relationship between II and K in the general case is given by the next lemma, which we will use later to bound sectional curvature in terms of bounds on the 2nd fundamental form.

Lemma 2.4.3. *For any 2-dimensional* $\sigma_w \subset T_w\mathcal{W}$*, let* $\nu^* \in T_w^\perp \mathcal{W}$ *be chosen to maximize* $|\det S_{\sigma_w,\nu}|$*. Then* $K(\sigma_w) = \det S_{\sigma_w,\nu^*}$*, hence* $|K_w| \leq \sup_{\{|u|=1\}} |\mathrm{II}_w(u, u)|^2$*.*

Proof. Assume $N - n \geq 2$ (otherwise the result is trivial). Fix an o.n. basis $\{u, v\}$ for

σ_w. For any o.n. basis $\{\nu^1, \dots, \nu^{N-n}\}$ of $T_w^\perp \mathcal{W}$, we have

$$
\begin{aligned}
K(\sigma_w) &= \sum_1^{N-n} \Big(\nu^k \cdot \mathrm{II}_w(u,u)\Big)\Big(\nu^k \cdot \mathrm{II}_w(v,v)\Big) - \Big(\nu^k \cdot \mathrm{II}_w(u,v)\Big)\Big(\nu^k \cdot \mathrm{II}_w(u,v)\Big) \\
&= \sum_1^{N-n} \det\Big(\nu^k \cdot \mathrm{II}_{\sigma_w}\Big) = \sum_1^{N-n} \det S_{\sigma_w,\nu^k}
\end{aligned}
$$

It is enough to demonstrate a choice of ν^k with $\det S_{\sigma_w,\nu^k} = 0$ for $k = 2, \dots, N-n$. If $\nu^k \perp \mathrm{Span}\{\mathrm{II}_w(u,u), \mathrm{II}_w(u,v)\}$ then it has no contribution to K, so we need only consider o.n. pairs ν^1, ν^2, chosen to span this subspace, and we need only show some $\nu = c_1\nu^1 + c_2\nu^2$ with $\alpha^2 + \beta^2 = 1$ and $\det S_{\sigma_w,\nu} = 0$. Let $S_k := S_{\sigma_w,\nu^k}$ for $k = 1, 2$. If $\det S_2 = 0$ we're done. Otherwise,

$$
\det S_{\sigma_w,\nu} = \det(c_1 S_1 + c_2 S_2) = c_1^2 (\det S_2) \det\left(S_1 S_2^{-1} + \frac{c_2}{c_1} I_2\right)
$$

The ratio c_2/c_1 may achieve any value in $\mathbb{R}$ while satisfying $c_1^2 + c_2^2 = 1$. Taking $c_2/c_1 = -\lambda$, where λ is an eigenvalue of $S_1 S_2^{-1}$, gives $\det S_{\sigma_w,\nu} = 0$, completing the proof. □

We also define two additional important measures of curvature which can be considered as ways to average the $K(\sigma_w)$ over various planes σ_w.

Definition 2.4.3. Let $\{u_1, \dots, u_n\}$ be any orthonormal basis for $T_w\mathcal{W}$. The *Ricci curvature* at w is a symmetric bilinear form on pairs $\tau_1, \tau_2 \in T_w\mathcal{W}$ given by

$$
\mathrm{Ric}_w(\tau_1, \tau_2) := \sum_{i=1}^{n} \langle \mathrm{II}_w(\tau_1,\tau_2), \mathrm{II}_w(u_i,u_i)\rangle - \langle \mathrm{II}_w(\tau_1,u_i), \mathrm{II}_w(\tau_2,u_i)\rangle
$$

and the *scalar curvature* at w is the real number

$$R_w := \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \mathrm{II}_w(u_i, u_i), \mathrm{II}_w(u_j, u_j) \rangle - \langle \mathrm{II}_w(u_i, u_j), \mathrm{II}_w(u_i, u_j) \rangle$$

Remark. The Ricci and scalar curvatures are independent of the choice of $\{u_1, \ldots, u_n\}$ and are completely determined by knowledge of the $K(\sigma_w)$, thus depend only on the intrinsic Riemannian metric g. Some authors take these as averages instead of sums, differing from our definitions by constant factors.

Curvature information can be used to describe the local relationship between volume and length. The following result is proven in [5, Ch. 3].

Theorem 2.4.4. *(Infinitesimal Bishop-Günther Inequalities) For $r > 0$, $\lambda \in \mathbb{R}$, define*

$$\psi(r) := \begin{cases} 1, & \lambda = 0 \\ \left(\lambda^{\frac{1}{2}} r\right)^{-1} \sin\left(\lambda^{\frac{1}{2}} r\right), & \lambda > 0 \\ \left(|\lambda|^{\frac{1}{2}} r\right)^{-1} \sinh\left(|\lambda|^{\frac{1}{2}} r\right), & \lambda < 0 \end{cases}$$

Let $\mathcal{W}$ be an n-dimensional smooth manifold, $n \geq 1$. Let $(r, \hat{\omega})$ be a polar geodesic normal coordinate system about $w_0 \in \mathcal{W}$, defined in $\mathcal{U} = B^{\mathcal{W}}_{\rho_\top}(w_0)$. We have $J_{w_0}(0) = 1$, and for $0 < r < \rho_\top$:

(a) If $K(\sigma_w) \geq \lambda$ for all σ_w, $w \in \mathcal{U}$, then $J_{w_0}(r\hat{\omega}) \leq [\psi(r)]^{n-1}$.

(b) If $K(\sigma_w) \leq \lambda$ for all σ_w, $w \in \mathcal{U}$, then $J_{w_0}(r\hat{\omega}) \geq [\psi(r)]^{n-1}$.

Remark. When $\lambda > 0$ and $r > \pi\lambda^{-1/2}$, Theorem 2.4.4 appears to give a negative upper bound for J_{w_0}, which is impossible. The conclusion is that when $\lambda > 0$, $\rho_\top \leq$

$\pi\lambda^{-1/2}$. This upper bound is achieved by the sphere of radius ρ, which has $\lambda = \rho^{-2}$ and $\rho_{\top} = \pi\rho$.

Corollary 2.4.5. *If $n \geq 1$ and $|\nu \cdot \mathrm{I\!I}_w(\tau,\tau)| \leq c_{\mathrm{I\!I}}|\nu||\tau|^2$ for every $\tau \in T_w\mathcal{W}$ and $\nu \in T_w^{\perp}\mathcal{W}$, then*

$$|J(r\hat{\omega}) - 1| \leq \left[\frac{\sinh(c_{\mathrm{I\!I}}r)}{c_{\mathrm{I\!I}}r}\right]^{n-1} - 1$$
$$\left|J(r\hat{\omega})^{-1} - 1\right| \leq 1 - \left[\frac{c_{\mathrm{I\!I}}r}{\sinh(c_{\mathrm{I\!I}}r)}\right]^{n-1}$$

Proof. Set $x := c_{\mathrm{I\!I}}r > 0$. By Lemma 2.4.3 and Theorem 2.4.4,

$$|J(r\hat{\omega}) - 1| \leq \left[\left(\frac{\sinh x}{x}\right)^{n-1} - 1\right] \vee \left[1 - \left(\frac{\sin x}{x}\right)^{n-1}\right]$$

From their power series expansions, $x > 0 \implies 1 < \frac{1}{2}\left(\frac{\sinh x}{x} + \frac{\sin x}{x}\right)$. By convexity of $x \mapsto x^{n-1}$, then, $1 < \left[\frac{1}{2}\left(\frac{\sinh x}{x} + \frac{\sin x}{x}\right)\right]^{n-1} \leq \frac{1}{2}\left[\left(\frac{\sinh x}{x}\right)^{n-1} + \left(\frac{\sin x}{x}\right)^{n-1}\right]$ which is equivalent to $\left(\frac{\sinh x}{x}\right)^{n-1} - 1 > 1 - \left(\frac{\sin x}{x}\right)^{n-1}$, yielding the first stated inequality. The second inequality follows similarly, by the concavity of $x \mapsto x^{1-n}$, with the inequality chain reversed. □

2.5 Tubular Neighborhoods

Definition 2.5.1 (Tubular Neighborhoods)**.**

(i) Let $\mathcal{W}$ be an n-dimensional differentiable submanifold of $\mathbb{R}^N$, $\rho > 0$, and $\mathcal{S}$ any subset of $\mathcal{W}$. Set the notation $B_\rho^{\perp}(\mathcal{S}) := \{w + \nu_w : w \in \mathcal{S}, \nu_w \in T_w^{\perp}\mathcal{W}, |\nu_w| < \rho\}$. We will also use $B_\rho^{\perp}(w) := B_\rho^{\perp}(\{w\})$ below.

(ii) If $\mathcal{U}$ is an open subset of $\mathcal{W}$ and every $y \in B_\rho^\perp(\mathcal{U})$ can be uniquely expressed in the form $y = w + \nu_w$ with $w \in \mathcal{U}, \nu_w \in T_w^\perp \mathcal{W}, |\nu_w| < \rho$, then $B_\rho^\perp(\mathcal{U})$ is called a *tubular neighborhood of radius ρ about $\mathcal{U}$.*

Theorem 2.5.1. *For any $w \in \mathcal{W}$ there is an open neighborhood $\mathcal{U} \subset \mathcal{W}$ containing w and a $\rho > 0$ such that $B_\rho^\perp(\mathcal{U})$ is a tubular neighborhood.*

This intuitively plausible result is proven in [9, pg. 200] in a more general setting.

Definition 2.5.2. Let $B_\rho^\perp(\mathcal{U})$ be a tubular neighborhood. At each $w \in \mathcal{U}$ the set $B_\rho^\perp(w) \subset T_w^\perp \mathcal{W}$ has the "obvious" $(N-n)$-dimensional Lebesgue measure dm^{N-n} induced by isometrically identifying it with $B_\rho^{N-n} \subset \mathbb{R}^{N-n}$. Define a Borel measure on $B_\rho^\perp(\mathcal{U})$ by the iterated integral

$$(V^n \times m^{N-n})(S) := \int_{\mathcal{W}} \int_{B_\rho^\perp(w)} 1_S(w + \nu_w)\, dm^{N-n}(\nu_w)\, dV^n(w)$$

Theorem 2.5.2. *Let $B_\rho^\perp(\mathcal{U})$ be a tubular neighborhood and dm^N the standard N-dimensional Lebesgue measure on it. Then*

$$\Theta(w, \nu) := \frac{dm^N}{dV^n dm^{N-n}} = \det(I_n - S_{w,\nu})$$

where I_n and $S_{w,\nu}$ are the identity operator and the shape operator on $T_w\mathcal{W}$, respectively.

This is proven in abstract modern notation in [5, ch. 3], and in classical notation, by Weyl, in [13].

Definition 2.5.3. Define *the even and odd parts of* Θ *with respect to* ν:

$$\Theta_e(w,\nu) := \frac{1}{2}(\Theta(w,\nu) + \Theta(w,-\nu))$$
$$\Theta_o(w,\nu) := \frac{1}{2}(\Theta(w,\nu) - \Theta(w,-\nu))$$

We will use these in subsequent sections, along with the following estimates:

Corollary 2.5.3. *If* $|\nu \cdot \mathrm{I\!I}_w(\tau,\tau)| \le c_{\mathrm{I\!I}}|\nu||\tau|^2$ *for every* $\tau \in T_w\mathcal{W}$, $\nu \in T_w^\perp\mathcal{W}$, *then*

$$(1 - c_{\mathrm{I\!I}}|\nu|)^n \le \Theta(w,\nu) \le (1 + c_{\mathrm{I\!I}}|\nu|)^n$$
$$|\Theta_e(w,\nu) - 1| \le \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{2k} (c_{\mathrm{I\!I}}|\nu|)^{2k}$$
$$|\Theta_o(w,\nu) - 1| \le \sum_{k=1}^{\lceil n/2 \rceil} \binom{n}{2k-1} (c_{\mathrm{I\!I}}|\nu|)^{2k-1}$$

Proof. Combining Theorems 2.4.1 and 2.5.2, we immediately have the first equation. For the other two, note that only even products of eigenvalues show up in the expansion of Θ_e via Theorem 2.5.2, and only odd products in Θ_o, then apply Theorem 2.4.1 to the sum. □

3

Entropy

This chapter defines and examines entropy with respect to a measure, with a focus on quantitative bounds on entropy differences for deducing convergence. Our ultimate focus will be entropy with respect to the induced volume measures of submanifolds of $\mathbb{R}^N$.

3.1 The Generalized Entropy Functional

Consider a random variable $X \in \mathbb{R}^N$ with probability density $p_X(x)$. The standard differential entropy is defined as $h(p_X) = -\int_{\mathbb{R}^N} p_X(x) \log p_X(x)\, dx$. With this definition, a differentiable and invertible change of variables $x \mapsto x'$ induces the "correction factor”

$$h(X') = \int p_{X'}(x') \log \frac{1}{p_{X'}(x')}\, dx' = h(X) + \mathbb{E}\left(\log\left|\frac{\partial x'}{\partial x}\right|\right)$$

It will be important in the work below to be able to work with differential entropy in a *coordinate-independent* manner, as manifolds generally cannot be parameterized entirely by any single coordinate system.

We adopt the following notation related to measure spaces and measures: $\mathcal{P}(\mathcal{M})$ is the set of positive measures on a measurable space $(\mathcal{M}, \Sigma)$, which we always take to be Borel (in fact, in applications $\mathcal{M}$ will always be a Borel subset of $\mathbb{R}^N$). The set of probability measures on $\mathcal{M}$ will be denoted by $\hat{\mathcal{P}}(\mathcal{M})$. If μ is a positive σ-finite measure on $(\mathcal{M}, \Sigma)$ then we will additionally take $\mathcal{P}(\mu) := \{f \in L^1(\mu)\colon f \geq 0\}$ and $\hat{\mathcal{P}}(\mu) := \{f \in L^1(\mu)\colon \|f\|_1 = 1\}$. Finally, we affirm the following minor abuse of notation: if P is a measure, then $P \in \hat{\mathcal{P}}(\mu)$ will indicate that P is a probability measure which is absolutely continuous with respect to μ (or equivalently, $\frac{dP}{d\mu} \in \hat{\mathcal{P}}(\mu)$.

Definition 3.1.1 (Generalized Entropy)**.** Let $(\mathcal{W}, \Sigma)$ be a measurable space with positive σ-finite measure μ, and let $f \in \mathcal{P}(\mu)$. The following quantities always exist as values in $[0, \infty]$:

$$h^+_\mu(f) := -\int_{0<f<1} f \log f \, d\mu$$
$$h^-_\mu(f) := \int_{f>1} f \log f \, d\mu$$

(i) If either $h^+_\mu(f)$ or $h^-_\mu(f)$ is finite, the *entropy of f with respect to* μ exists as a value in $[-\infty, \infty]$ and is defined by

$$h_\mu(f) := h^+_\mu(f) - h^-_\mu(f) = -\int_{\{f>0\}} f \log f \, d\mu$$

If $h^+_\mu(f) = h^-_\mu(f) = \infty$, $h_\mu(f)$ does not exist (is undefined).

(ii) For $P \in \hat{\mathcal{P}}(\mu)$, $h^\pm_\mu(P) := h^\pm_\mu(\frac{dP}{d\mu})$. The *entropy of P with respect to μ* is $h_\mu(P) := h_\mu(\frac{dP}{d\mu}) \in [-\infty, \infty]$ whenever $h_\mu(\frac{dP}{d\mu})$ exists.

Remark. When $h_\mu(P)$ exists,

$$h_\mu(P) = -\mathbb{E}\log\frac{dP}{d\mu} = -\int_{\mathcal{W}} \log\frac{dP}{d\mu}\, dP = -\int_{\frac{dP}{d\mu}>0} \frac{dP}{d\mu}\log\frac{dP}{d\mu}\, d\mu$$

$h_\mu(P)$ generalizes the two classical entropies of information theory: If $\mathcal{W} = \mathbb{R}^N$, Σ is Borel sets, and $d\mu = dm = dx^1 dx^2 \cdots dx^N$, the standard N-dimensional Lebesgue measure, then $h_m(P) = h(P)$, the standard differential entropy on $\mathbb{R}^N$. If $\mathcal{W} = \mathbb{Z}^+$ and μ is the counting measure, then $\frac{dP}{d\mu}(i) = P(i) = p_i$, and $h_\mu(P) = -\sum p_i \log p_i$ is the standard discrete entropy on countable spaces.

Many classical properties of $h(P)$ proven by convexity can be extended to arbitrary μ. One example is the following, which will be of use later:

Lemma 3.1.1. *Let μ be a positive measure on $\mathcal{W}$ with $\mu(\mathcal{W}) < \infty$. Then $h_\mu(P)$ exists for every $P \in \hat{\mathcal{P}}(\mu)$, $\sup_P h_\mu(P) = \log\mu(\mathcal{W})$, and the sup is achieved by the constant density $p = \mu(\mathcal{W})^{-1} d\mu$.*

Proof. $\int \mu(\mathcal{W})^{-1}\log\mu(\mathcal{W})\, d\mu = \log\mu(\mathcal{W})$, so the constant density achieves the stated entropy. Now let $P \in \hat{\mathcal{P}}(\mu)$, set $p = \frac{dP}{d\mu}$, and $\hat{\mu} := \mu(\mathcal{W})^{-1}\mu$, and note that $\int (\mu(\mathcal{W})p)\, d\hat{\mu} = \int p\, d\mu = 1$. Applying Jensen's inequality with the convex function $x \mapsto x\log x$ on

$x > 0$, we have

$$
\begin{aligned}
0 = 1 \log 1 &= \left[\int (\mu(\mathcal{W})p) \, d\hat{\mu} \right] \log \left[\int (\mu(\mathcal{W})p) \, d\hat{\mu} \right] \\
&\leq \int_{p>0} (\mu(\mathcal{W})p) \log(\mu(\mathcal{W})p) \, d\hat{\mu} = \int_{p>0} p \log(\mu(\mathcal{W})p) \, d\mu
\end{aligned}
$$

Thus $\log \mu(\mathcal{W}) \geq \log \mu(\mathcal{W}) - \int_{p>0} p \log(\mu(\mathcal{W})p) \, d\mu = -\int_{p>0} p \log p \, d\mu = h_\mu(p)$, which shows existence and the stated bound. $\square$

3.2 General Entropy Estimates

The following results will be needed to prove uniform convergence of approximate entropies. Note that here and elsewhere we will make use of the binary maximum and minimum operators: $a \vee b := \max\{a, b\}$, $a \wedge b := \min\{a, b\}$.

Lemma 3.2.1. *For $a, b \geq 0$ and $p > 0$ we have $(a + b)^p \leq (2^{p-1} \vee 1)(a^p + b^p)$ and $a^p + b^p \leq (2^{1-p} \vee 1)(a + b)^p$.*

Proof. This result is standard and is easily proven with calculus; Extremize $\psi(x) = (x + b)^p (x^p + b^p)^{-1}$ for $x \in [0, \infty)$. $\square$

Lemma 3.2.2. *The function defined for $t \geq 0$ by $\psi(0) = 0$, $\psi(t) = t \log \frac{1}{t}$ when $t > 0$ is uniformly Hölder continuous on $[0, A]$ for any exponent $\alpha \in (0, 1)$ and $A > 0$. In particular, if $0 \leq s < t \leq A$ then*

$$
\frac{|\psi(t) - \psi(s)|}{|t - s|^\alpha} \leq \frac{1}{1 - \alpha} \vee A^{1-\alpha} \log^+(eA)
$$

Proof. Note that ψ is continuous for $t \geq 0$, smooth for $t > 0$, increasing for $0 \leq t < e^{-1}$, decreasing for $t > e^{-1}$, and concave down. We now consider several cases. If

$s = 0 < t \le A \le e^{-1}$ then we have

$$\begin{aligned}\frac{|\psi(t) - \psi(s)|}{|t-s|^\alpha} &= t^{1-\alpha} \log \frac{1}{t} \\ &\le c^0_{\alpha,e^{-1}} := \sup_{0 \le t \le e^{-1}} t^{1-\alpha} |\log t|\end{aligned}$$

$t^{1-\alpha}\log t$ is non-negative and continuous on $[0, e^{-1}]$ and achieves its only stationary point at $t = e^{-\frac{1}{1-\alpha}}$, so $c^0_{\alpha,e^{-1}} = \frac{1}{e(1-\alpha)}$.

If $0 < s < t \le A \le e^{-1}$, since ψ is increasing we need only bound from above $[\psi(t) - \psi(s)](t-s)^{-\alpha}$, and since ψ is concave-down,

$$\begin{aligned}\psi(t) &\le \psi(s) + \psi'(s)(t-s) \\ &\le \psi(s) + \left(\log \frac{1}{es}\right)(t-s) \\ &\le \psi(s) + (t-s)^{1-\alpha}\left(\log \frac{1}{es}\right)(t-s)^\alpha\end{aligned}$$

If we also have $s \ge e^{-1}(t-s)$ then, using $t-s \in (0, A]$, we have $[\psi(t) - \psi(s)](t-s)^{-\alpha} \le \left[(t-s)^{1-\alpha} \log \frac{1}{t-s}\right] \le c^0_{\alpha,e^{-1}}$. If we have instead $s < e^{-1}(t-s)$, then $t = s + (t-s) < (1+e^{-1})(t-s)$, so $\psi(t) - \psi(s) \le \psi(t) \le \psi\big((1+e^{-1})(t-s)\big) \le c^0_{\alpha,e^{-1}}\big(1+e^{-1}\big)^\alpha (t-s)^\alpha \le (1+e^{-1})c^0_{\alpha,e^{-1}}(t-s)^\alpha$.

In the case $e^{-1} \le s < t \le A$ we have, by the Mean Value Theorem, $|\psi(t) - \psi(s)| \le (\log eA)(t-s) \le \big[A^{1-\alpha}\log(eA)\big](t-s)^\alpha$.

Finally, if $0 < s < e^{-1} < t \le A$ then

$$|\psi(t) - \psi(s)| \le \big|\psi(t) - \psi(e^{-1})\big| \vee \big|\psi(s) - \psi(e^{-1})\big|$$

Since $(t-e^{-1})^\alpha, (e^{-1}-s)^\alpha \le (t-s)^\alpha$ we can always apply one of the previous cases to achieve a bound. $\square$

Theorem 3.2.3. *Let $\alpha \in (0,1)$, $b \in (1,\infty]$, and $f, g \in \mathcal{P}(\mu)$ for measure μ. Put $S := \{f \vee g > 1\}$. Suppose $h_\mu(g)$ exists, and $\|g\|_{1;S}$, $|h_\mu(g)|$, $\int |f-g|^\alpha \, d\mu$, and $\|f-g\|_{b;S}$ are all finite. Then $h_\mu(f)$ exists, is finite, and*

$$|h_\mu(f) - h_\mu(g)| \le \frac{1}{1-\alpha} \int |f-g|^\alpha \, d\mu + \frac{e^{\alpha(1-b^{-1})} \|f+g\|_{1;S}^{1-\alpha b^{-1}}}{\alpha(1-b^{-1})} \|f-g\|_{b;S}^\alpha$$

Proof. Note that $\|f+g\|_{1;S} \le 2\|g\|_{1;S} + \|f-g\|_{1;S} \le 2\|g\|_{1;S} + \|f-g\|_{b;S}$, so $\|f+g\|_{1;S} < \infty$. Put f, g in place of s, t in the previous lemma. On S^c, the bound is simply $|\psi(f) - \psi(g)| \le (1-\alpha)^{-1} \vee 1 \le (1-\alpha)^{-1}$. On S, A may be set to $f+g$ and the the maximum bound replaced by a sum, giving $|\psi(f) - \psi(g)| \le (1-\alpha)^{-1} + (f+g)^{1-\alpha} \log^+[e(f+g)]$. Define $\gamma = \alpha(1-b^{-1}) \in (0, \alpha]$. On S, $[e(f+g)]^\gamma > 1$, so $\log^+[e(f+g)] = \gamma^{-1} \log^+[e(f+g)]^\gamma \le e^\gamma \gamma^{-1} (f+g)^\gamma$. Integrate these bounds and apply Hölder's inequality with the exponents $(1-\alpha+\gamma)^{-1}$ and $(\alpha-\gamma)^{-1} = \alpha^{-1} b$:

$$\begin{aligned}
|h_\mu(f) - h_\mu(g)| &\le \int_{S^c} \frac{1}{1-\alpha} |f-g|^\alpha \, d\mu + \\
&\qquad + \int_S \left(\frac{1}{1-\alpha} + (f+g)^{1-\alpha} \log^+[e(f+g)] \right) |f-g|^\alpha \, d\mu \\
&\le \frac{1}{1-\alpha} \int |f-g|^\alpha \, d\mu + \gamma^{-1} e^\gamma \int_S (f+g)^{1-\alpha+\gamma} |f-g|^\alpha \, d\mu \\
&\le \frac{1}{1-\alpha} \int |f-g|^\alpha \, d\mu + \gamma^{-1} e^\gamma \|f+g\|_{1;S}^{1-\alpha b^{-1}} \|f-g\|_{b;S}^\alpha
\end{aligned}$$

$\square$

Theorem 3.2.4. *Let $\alpha \in (0,1)$, $b \in (1,\infty]$, and $f, g \in \mathcal{P}(\mu)$ for measure μ. Define $c_\pm := \left| \int f \pm g \, d\mu \right|$, and suppose $h_\mu(g)$ exists and is finite, $\int |f-g|^\alpha \, d\mu \le c_\alpha < \infty$, and*

$\|f-g\|_b \leq c_b < \infty$. *Then $h_\mu(f)$ exists, is finite, and*

$$|h_\mu(f) - h_\mu(g)| \leq \frac{c_+^{1-\alpha} e^{\frac{\alpha(1-\alpha)(1-b^{-1})}{1-\alpha b^{-1}}} (1-\alpha b^{-1})}{\alpha(1-\alpha)(1-b^{-1})} \left({c_\alpha}^{1-b^{-1}} {c_b}^{1-\alpha} \right)^{\frac{\alpha}{1-\alpha b^{-1}}}$$
$$+ c_- \left| \log(c_+) + \frac{1}{1-\alpha b^{-1}} \left(\alpha(1-b^{-1}) + \alpha \log c_b - \log c_\alpha \right) \right|$$

In the important case when $\|f\|_1 = \|g\|_1 = 1$ (when f, g are probability distributions), we have

$$|h_\mu(f) - h_\mu(g)| \leq \frac{2^{1-\alpha} e^{\frac{\alpha(1-\alpha)(1-b^{-1})}{1-\alpha b^{-1}}} (1-\alpha b^{-1})}{\alpha(1-\alpha)(1-b^{-1})} \left({c_\alpha}^{1-b^{-1}} {c_b}^{1-\alpha} \right)^{\frac{\alpha}{1-\alpha b^{-1}}}$$

Proof. By the log-convexity of the L^p norms, $\|u\|_1 \leq \left(\int |u|^\alpha \, d\mu\right)^{\frac{1-b^{-1}}{1-\alpha b^{-1}}} \|u\|_b^{\frac{1-\alpha}{1-\alpha b^{-1}}}$, so $c_\alpha, c_b < \infty \implies c_\pm < \infty$. For every constant scale factor $r > 0$, we can compute

$$h_\mu(f) - h_\mu(g) = h_{r\mu}(r^{-1} f) - h_{r\mu}(r^{-1} g) + (\log r) \int f - g \, d\mu$$

Theorem 3.2.3 gives, for all $r > 0$, $b < \infty$,

$$\left| h_{r\mu}\left(\frac{f}{r}\right) - h_{r\mu}\left(\frac{g}{r}\right) \right| \leq \frac{1}{1-\alpha} \int \left| \frac{f-g}{r} \right|^\alpha r \, d\mu + \frac{c_+^{1-\alpha b^{-1}} e^{\alpha(1-b^{-1})}}{\alpha(1-b^{-1})} \left(\int \left| \frac{f-g}{r} \right|^b r \, d\mu \right)^{\alpha b^{-1}}$$
$$\leq \frac{r^{1-\alpha}}{1-\alpha} \int |f-g|^\alpha \, d\mu + \frac{c_+^{1-\alpha b^{-1}} e^{\alpha(1-b^{-1})} r^{-\alpha(1-b^{-1})}}{\alpha(1-b^{-1})} \|f-g\|_b^\alpha$$

where we have simplified by replacing S with the entire measure space. It is easily checked that the same inequality holds for $b = \infty$. Thus, we have,

$$|h_\mu(f) - h_\mu(g)| \leq \frac{c_\alpha}{1-\alpha} r^{1-\alpha} + \frac{c_+^{1-\alpha b^{-1}} e^{\alpha(1-b^{-1})} c_b^\alpha}{\alpha(1-b^{-1})} r^{-\alpha(1-b^{-1})} + c_1 |\log r|$$

for all $r > 0$. Taking $r = \left(c_+^{1-\alpha b^{-1}} e^{\alpha(1-b^{-1})} c_b^{\alpha} c_\alpha^{-1}\right)^{\frac{1}{1-\alpha b^{-1}}}$ gives the stated bound. If $\|f\|_1 = \|g\|_1$ then $c_+ = 2$ and $c_- = 0$, giving the second assertion. □

Thus the rate of convergence of the differential entropies of a sequence may be estimated in terms of the L^α quasi-norm and L^b norm of the differences. In general, neither of these two quantities alone is sufficient to deduce convergence.

We can use these quantities to deduce existence and finiteness properties of h_μ:

Corollary 3.2.5. *Let $f \in \mathcal{P}(\mu), p \in \hat{\mathcal{P}}(\mu)$. Each of the following imply that $h_\mu(f)$ exists, and satisfies the stated upper or lower bound, whenever the quantities in the bound are finite:*

(i) For any $\alpha \in (0,1)$: $h_\mu(f) \le \frac{1}{1-\alpha}\int_{f<1} f^\alpha \, d\mu$.

(ii) $h_\mu(f) \le \mu(\{0 < f < 1\})$.

(iii) For $b \in (1,\infty)$: $h_\mu(f) > -2\big[1 + e(1-b^{-1})^{-1}\big]\|f\|_{b;\{f>1\}}^b$.

(iv) For $b \in (1,\infty]$: $h_\mu(p) > -\big[1 + e^2(1-b^{-1})^{-1}\big]\log\Big(\|p\|_{b;\{p>1\}} \vee e^2\Big)$.

Proof. Set $f^+ := 1_{\{f<1\}}f$, so $h_\mu(f^+) = h_\mu^+(f) \ge 0$. Applying Theorem 3.2.3 with f^+ in place of f and $g = 0$ gives $h_\mu^+(f) \le (1-\alpha)^{-1}\int_{0<f<1} f^\alpha \, d\mu$. This gives (i) immediately, and (ii) follows from (i) since $\int_{\{0<f<1\}} f^\alpha \, d\mu \le \mu(\{0 < f < 1\})$ for all $\alpha \in (0,1)$.

Now set $f^- := 1_{\{f>1\}}f$, so $h_\mu(f^-) = -h_\mu^-(f) \le 0$. For any $\alpha \in (0,1)$ we have $\int_{\{f>1\}} f^\alpha \, d\mu < \|f\|_{1;\{f>1\}}$. We can apply Theorem 3.2.3 for any $\alpha \in (0,1)$, f^- in place of f, and $g = 0$ to obtain that, whenever the quantities on the RHS are finite,

$$h_\mu^-(f) \le \frac{1}{1-\alpha}\|f\|_{1;\{f>1\}} + \frac{e}{\alpha(1-b^{-1})}\|f\|_{1;\{f>1\}}^{1-\alpha b^{-1}}\|f\|_{b;\{f>1\}}^{\alpha}$$

(iii) follows by taking $\alpha = \frac{1}{2}$ and noting that $\|f\|_{1;\{f>1\}} \leq \|f\|_{b;\{f>1\}}^b$. For (iv), take $\alpha = \left[\log\left(\|f\|_{b;\{f>1\}} \vee e^2\right)\right]^{-1}$, so $\alpha \leq \frac{1}{2} \implies (1-\alpha)^{-1} \leq \alpha^{-1}$. After some algebra, the stated bound follows. $\square$

3.3 Renormalization Entropy Estimate

This simple lemma lets us estimate certain entropy errors due to an overall scaling factor u which will be close to unity.

Lemma 3.3.1. *Suppose $p \in \hat{\mathcal{P}}(\mu)$, $h_\mu(p)$ is finite, and $u \in L^\infty(\mu)$ with $u \geq 0$.*

(i) In the special case when $u \equiv u_0$ is constant μ-a.e.: $h_\mu(up)$ exists, and

$$|h_\mu(up) - h_\mu(p)| \leq |u_0 - 1|[(1 + u_0) + |h_\mu(p)|]$$

(ii) If instead $p \in L^\infty(\mu)$, then $h_\mu(up)$ exists, and

$$|h_\mu(up) - h_\mu(p)| \leq \|u-1\|_\infty(1 + \|u\|_\infty + 2\|p\|_\infty + h_\mu(p))$$

Proof. We have

$$\begin{aligned} |h_\mu(up) - h_\mu(p)| &= \left|\int up\log(up) - p\log p\, d\mu\right| \\ &\leq \left|\int (u\log u)p\, d\mu\right| + \left|\int (u-1)p\log p\, d\mu\right| \end{aligned}$$

The first term is bounded by $\|u\log u\|_\infty$. Since $x\log x \leq x(x-1)$ when $x > 1$ and $x - 1 \leq x\log x$ when $0 < x < 1$, we always have $\|u\log u\|_\infty \leq \|u-1\|_\infty(1 + \|u\|_\infty)$.

If u is constant then $\left|\int (u-1)p\log p\, d\mu\right| = |u_0 - 1||h_\mu(p)|$, and (i) is proven.

Otherwise, for (ii), $\left|\int (u-1) p \log p \, d\mu\right| \leq \|u-1\|_\infty \int p |\log p| \, d\mu$. Let $S = \{p > 1\}$, so $\int p|\log p| \, d\mu = h_\mu(p) + 2\int_S p \log p \, d\mu$. To complete the proof, note that $\int_S p \log p \, d\mu \leq \int_S p(p-1) \, d\mu \leq \|(p-1)^+\|_\infty \leq \|p\|_\infty$. □

3.4 Entropy Estimates on Uniform Submanifolds of $\mathbb{R}^N$

Let $\mathcal{W}$ be a differentiable submanifold of $\mathbb{R}^N$ of dimension n, with $N > 0$ and $0 \leq n \leq N$. We will sometimes refer to the codimension as $n' := N - n$. We specifically wish to allow the possibility that the closure $\overline{\mathcal{W}}$ is a manifold-with-boundary.

The standard Euclidean metric on $\mathbb{R}^N$ induces a metric on $\mathcal{W}$, which in turn induces a volume measure supported on $\mathcal{W}$ which we denote V^n (the "n-dimensional volume" of sets in $\mathcal{W}$). When P_X is a probability measure on $\mathcal{W}$ satisfying $V^n(E) = 0 \implies P_X(E) = 0$, the probability density p_X exists as the Radon-Nikodym derivative $\frac{dP_X}{dV^n}$. The entropy of such P_X will be computed relative to V^n:

$$h_{V^n}(p_W) = \int_{\mathcal{W}} p_W(w) \log \frac{1}{p_W(w)} \, dV^n$$

In this and the following chapter we will utilize several convenient uniformity assumptions. We require the existence of a tubular neighborhood $\mathcal{U} \subset \mathbb{R}^N$ about $\mathcal{W}$ which is of uniform radius $\rho_\perp(\mathcal{W}) > 0$. We will require that the 2nd fundamental form of $\mathcal{W} \subset \mathbb{R}^N$ is uniformly bounded.* That is, we assume

$$c_{\mathrm{II}}(\mathcal{W}) := \sup_{w \in \mathcal{W},\, \tau \in T_w\mathcal{W},\, |\tau|=1} |\mathrm{II}_w(\tau, \tau)| < \infty$$

These two requirements suffice for the results of this chapter, but we add the following additional requirement that will be needed in the next chapter: If $\mathcal{W}$ has no boundary, we require a $\rho_{\top}(\mathcal{W}) > 0$ such that a geodesic normal coordinate system of geodesic radius $\rho_{\top}(\mathcal{W})$ can be constructed at all $w \in \mathcal{W}$. If $\partial\mathcal{W} \neq \emptyset$ we instead require that there exists an n-dimensional submanifold $\mathcal{W}' \supset \mathcal{W}$, and a $\rho_{\top}(\mathcal{W}) > 0$ such that a geodesic normal coordinate system for $\mathcal{W}'$, of geodesic radius $\rho_{\top}(\mathcal{W})$, can be constructed at all $w \in \mathcal{W}$.

For notational and computational convenience, we collect the above into two bounding constants:

Definition 3.4.1. If $\mathcal{W} \subset \mathbb{R}^N$ is a differentiable submanifold and $\rho_{\perp}(\mathcal{W}), \rho_{\top}(\mathcal{W})$, and $c_{\mathrm{II}}(\mathcal{W})$ are as described above, define

$$c^o_{\mathcal{W}} := \max\bigl\{c_{\mathrm{II}}(\mathcal{W}), \rho_{\perp}(\mathcal{W})^{-1}\bigr\}$$
$$c_{\mathcal{W}} := \max\bigl\{c_{\mathrm{II}}(\mathcal{W}), \rho_{\perp}(\mathcal{W})^{-1}, \rho_{\top}(\mathcal{W})^{-1}\bigr\}$$

If $c^o_{\mathcal{W}} < \infty$ we will say that $\mathcal{W}$ is *semi-uniform*. If $c_{\mathcal{W}} < \infty$ we will say that $\mathcal{W}$ is *uniform*.

*This requirement is a way to control the degree to which the geometry of $\mathcal{W}$ deviates from the standard Euclidean geometry. Here is an equivalent requirement: At a point $w \in \mathcal{W}$ we can choose a direction w tangent to $\mathcal{W}$ and a direction ν normal to $\mathcal{W}$, which together span a plane P. In a small region of P near w, $\mathcal{W} \cap P$ can be "best approximated" by an arc of a circle of some radius which is tangent to $\mathcal{W} \cap P$ at w (where radius of ∞ is permitted in the case of a straight line). Our requirement is that the radius of such a circle is always $\geq c_{\mathrm{II}}^{-1}$ for every w, τ, and ν.

Remark.

(i) If $\mathcal{W}$ is compact, $c_{\mathcal{W}} < \infty$.

(ii) $c_{\mathcal{W}}$ scales like length^{-1}, so if $a > 0$ and $a\mathcal{W} = \{aw \colon w \in \mathcal{W}\}$, $c_{a\mathcal{W}} = a^{-1}c_{\mathcal{W}}$.

(iii) $c_{\mathcal{W}} = 0 \implies \mathcal{W}$ is an affine linear subspace of $\mathbb{R}^N$, that is, a copy of $\mathbb{R}^n$ up to some fixed rotation and translation.

(iv) Any open subset of $\mathbb{R}^N$ is semi-uniform, but need not be uniform. We will need this distinction later.

For brevity, we use $\kappa_{N,n} := \frac{N^N\Gamma(1+n/2)\Gamma(1+n'/2)}{n^n(n')^{n'}\Gamma(1+N/2)}$ in the following results, following the convention $0^0 := 1$ when needed. It can be verified, with the Sterling approximation bounds, that $\kappa_{N,n} \leq 2N^{1/2}e^{N/2}$ for all $n, N \geq 0$.

The following lemma allows us to bound dV^n-integration on $\mathcal{W}$ using standard Lebesgue integration on $\mathbb{R}^N$ (expressed here in spherical coordinates). If the assumption of $\rho_\perp > 0$ is omitted, the lemma no longer holds, with counterexamples provided by "space-filling" curves.

Lemma 3.4.1. *Let $\mathcal{W}$ be semi-uniform and define $V^n_{\mathcal{W}}(r) := V^n(\mathcal{W} \cap B^N_r)$. We have*

$$V^n_{\mathcal{W}}(r) \leq \kappa_{N,n}\, \omega_n r^n (1 + c^o_{\mathcal{W}} r)^{n'}$$

Suppose $\psi \in L^1[r_0, \infty)$ is differentiable with $\psi' \in L^1[r_0, \infty)$, $\psi \geq 0$, and $\lim_{r\to\infty} \psi(r) r^N = 0$. Then we have

$$\int_{\mathcal{W}\cap\{|w|>r_0\}} \psi(|w|)\, dV^n \leq \kappa_{N,n}\, \omega_n \int_{\{r>r_0\colon\, \psi'(r)<0\}} \left|\psi'(r)\right| r^n (1 + c^o_{\mathcal{W}} r)^{n'}\, dr$$

Proof. Let $\rho > 0$. We have $V^N(B_\rho(\mathcal{W} \cap B_r^N)) \le \omega_N (r+\rho)^N$ by the triangle inequality. On the other hand, when $\rho < \rho_\perp(\mathcal{W})$, this is bounded below by integrating in tubular coordinates and applying Corollary 2.5.3:

$$\begin{aligned} V^N(B_\rho(\mathcal{W} \cap B_r^N)) &= \int_{\mathcal{W}\cap B_r^N} \int_{B_\rho^{n'}(w)} \Theta(w,\nu)\, d\nu^{n'}\, dV^n \\ &\ge V^n(\mathcal{W} \cap B_r^N)\omega_{n'}\rho^{n'}(1 - c_{\mathrm{II}}(\mathcal{W})\rho)^n \end{aligned}$$

Chaining the inequalities and using $c_{\mathrm{II}}(\mathcal{W}) \le c_{\mathcal{W}}^o$ gives

$$V_{\mathcal{W}}^n(r) \le \frac{\omega_N (r+\rho)^N}{\omega_{n'}\rho^{n'}(1 - c_{\mathcal{W}}^o \rho)^n}$$

whenever $c_{\mathcal{W}}^o \rho < 1$. Setting $\rho = (n'r)(n + Nrc_{\mathcal{W}}^o)^{-1}$ (the minimizing choice) yields the first claim, after some algebra.

$V_{\mathcal{W}}^n(r)$ is an increasing lower-semicontinuous function of $r \ge 0$ (possibly containing a countable set of jump discontinuities, which occur at values of r for which $S^{N-1}(r)$ contains a non-empty n-dimensional submanifold of $\mathcal{W}$). Thus it has a generalized (distributional) derivative with respect to r which obeys the generalized integration-by-parts formula:

$$\begin{aligned} \int_{\mathcal{W}\cap\{r_0<|w|\le r_1\}} \psi(|w|)\, dV^n &= \int_{(r_0,r_1]} \psi(|r|) \frac{dV_{\mathcal{W}}^n(r)}{dr}\, dr \\ &= [\psi(r) V_{\mathcal{W}}^n(r)]_{r_0}^{r_1} - \int_{(r_0,r_1]} \psi'(r) V_{\mathcal{W}}^n(r)\, dr \end{aligned}$$

Applying the previous bounds and taking the limit as $r_1 \to \infty$ yields the assertion.

□

The following weighted norms provide a way to quantify how quickly a function goes to zero in the large-radius limit.

Definition 3.4.2. Let $\mathcal{W}$ be semi-uniform and $\delta \geq 0$. For V^n-measurable $f\colon \mathcal{W} \to \mathbb{R}$, the *decay norm of exponent* δ is

$$\|f\|_{(\delta)} := \int_{\mathcal{W}} |f(w)|(1 + c^o_{\mathcal{W}}|w|)^\delta \, dV^n$$

If P_W is a probability measure on $\mathcal{W}$,

$$\|P\|_{(\delta)} := \mathbb{E}(1 + c^o_{\mathcal{W}}|W|)^\delta$$

More generally, for any $a_{\mathcal{W}} \geq c^o_{\mathcal{W}}$ we define

$$\|f\|_{(\delta);a_{\mathcal{W}}} := \int_{\mathcal{W}} |f(w)|(1 + a_{\mathcal{W}}|w|)^\delta \, dV^n$$

$$\|P_W\|_{(\delta);a_{\mathcal{W}}} := \mathbb{E}(1 + a_{\mathcal{W}}|W|)^\delta$$

Remark.

(i) $\|\cdot\|_{(0)} = \|\cdot\|_{L^1(V^n)}$, and $\|\cdot\|_{(\delta)}$ is monotonically increasing in δ.

(ii) If $\frac{dP_W}{dV^n}$ exists, $\|P_W\|_{(\delta);a_{\mathcal{W}}} = \left\|\frac{dP_W}{dV^n}\right\|_{(\delta);a_{\mathcal{W}}}$.

We always have $\|P_W\|_{(\delta);a_{\mathcal{W}}} \leq (2^{\delta-1} \vee 1)(1 + a^\delta_{\mathcal{W}}\, \mathbb{E}|W|^\delta)$. If $\delta \leq 2$, Jensen's inequality then gives $\|P_W\|_{(\delta);a_{\mathcal{W}}} \leq (2^{\delta-1} \vee 1)\left(1 + a^\delta_{\mathcal{W}} \left[\mathbb{E}|W|^2\right]^{\delta/2}\right)$. So an average power constraint on W implies a decay norm bound on its probability distribution for all $0 < \delta \leq 2$. The next theorem shows that decay norm bounds imply L^α bounds (which, by Theorem 3.2.3, are the key component in entropy estimates):

Lemma 3.4.2. *Let $\mathcal{W}$ be semi-uniform with $a_{\mathcal{W}} \geq c^o_{\mathcal{W}}$, and $f : \mathcal{W} \to \mathbb{R}$ be V^n-measurable. Suppose $\delta > 0$, and $\frac{N+\delta}{N+2\delta} \leq \alpha < 1$. Then,*

$$\int_{\mathcal{W}} |f(w)|^\alpha \, dV^n \leq e \left(\frac{\kappa_{N,n}\omega_n}{a^n_{\mathcal{W}}} \right)^{1-\alpha} \|f\|^\alpha_{(\delta);a_{\mathcal{W}}}$$

Proof. We first consider the boundary case $\delta = \frac{N(1-\alpha)}{2\alpha-1}$, which is equivalent to $\alpha = \frac{N+\delta}{N+2\delta}$. Note also that $\alpha\delta = (1-\delta)(N+\delta)$.

Set $\eta(r) := 1 + a_{\mathcal{W}}|w|$. Multiply $|f|^\alpha$ by $1 = \eta^{\delta\alpha}\eta^{-\delta\alpha}$ and apply Hölder's inequality with conjugate exponents $(1-\alpha)^{-1}$ and α^{-1}:

$$\begin{aligned} \int_{\mathcal{W}} |f(w)|^\alpha \, dV^n &= \int_{\mathcal{W}} \eta(|w|)^{-\delta\alpha} \Big[|f(w)| \eta(|w|)^\delta \Big]^\alpha \, dV^n \\ &\leq \left[\int_{\mathcal{W}} \eta(|w|)^{-(N+\delta)} \, dV^n \right]^{1-\alpha} \left[\int_{\mathcal{W}} f(w)\eta(|w|)^\delta \right]^\alpha \end{aligned}$$

Apply Lemma 3.4.1 to the first bracketed term and use $a_{\mathcal{W}} \geq c^o_{\mathcal{W}}$:

$$\begin{aligned} \int_{\mathcal{W}} \eta(|w|)^{-(N+\delta)} \, dV^n &\leq \kappa_{N,n}\omega_n (N+\delta) a_{\mathcal{W}} \int_0^\infty \left(\frac{r}{1+a_{\mathcal{W}}r} \right)^n \left(\frac{1+c^o_{\mathcal{W}}r}{1+a_{\mathcal{W}}r} \right)^{n'} \eta(r)^{-1-\delta} \, dr \\ &\leq \kappa_{N,n}\omega_n (N+\delta) a_{\mathcal{W}}{}^{1-n} \int_0^\infty (1+a_{\mathcal{W}}r)^{-1-\delta} \, dr \\ &\leq \kappa_{N,n}\omega_n (1+N\delta^{-1}) a_{\mathcal{W}}{}^{-n} \end{aligned}$$

Combining the inequalities and recognizing that $(1+N\delta)^{1-\alpha} = (1+N\delta^{-1})^{\delta/(N+2\delta)} \leq (1+N\delta^{-1})^{\delta/N} \leq e$ completes the proof of the boundary case. The general case then follows by the monotonicity of $\|\cdot\|_{(\delta);a_{\mathcal{W}}}$ with respect to δ. $\square$

Corollary 3.4.3. *Let $\mathcal{W}$ be semi-uniform with $a_{\mathcal{W}} \geq c^o_{\mathcal{W}}$, $\delta > 0$.*

(i) If $f \in \mathcal{P}(V^n)$ and $\|f\|_{(\delta);a_{\mathcal{W}}} < \infty$, then $h_{V^n}(f)$ exists, and

$$h_{V^n}(f) \le e^3 \left[\left(2 + N\delta^{-1}\right) \vee \log\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right) \vee \log\|f\|_1^{-1} \right] \|f\|_{(\delta);a_{\mathcal{W}}}$$

(ii) Every $P_W \in \hat{\mathcal{P}}(V^n)$ with $\mathbb{E}|W|^\delta \le \mathsf{K}$ has well-defined $h_{V^n}(P_W)$ bounded by

$$h_{V^n}(P_W) \le (2^{\delta-1} \vee 1) e^3 \left[\left(2 + N\delta^{-1}\right) \vee \log\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right) \right] (1 + a_{\mathcal{W}}^\delta \mathsf{K})$$

Proof. Combine Corollary 3.2.5(i) with Lemma 3.4.2 and take

$$\alpha = 1 - \left[\left(2 + N\delta^{-1}\right) \vee \log\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right) \vee \log\|f\|_{(\delta);a_{\mathcal{W}}}^{-1} \right]^{-1} \ge \frac{N+\delta}{N+2\delta}$$

Since $\|f\|_1 \le \|f\|_{(\delta);a_{\mathcal{W}}}$, this gives (i). For (ii), use $\mathbb{E}(1 + a_{\mathcal{W}}|W|)^\delta \le (2^{\delta-1} \vee 1)(1 + a_{\mathcal{W}}^\delta \mathbb{E}|W|^\delta)$. □

Remark. In particular, this corollary proves that all $P_W \ll V^n$ satisfying an average power constraint $\mathbb{E}|W|^2 \le \mathsf{P}$ have a well-defined entropy that is bounded above in terms of the constant P.

We now combine our previous results into our primary tools for proving entropy estimates and convergence results.

Theorem 3.4.4. *Let $\mathcal{W}$ be semi-uniform with $a_{\mathcal{W}} \ge c_{\mathcal{W}}^o$. Let $\delta > 0$, $b \in (1, \infty]$. Suppose $f, g \in L_+^1(V^n)$, $h_{V^n}(g)$ exists and is finite, $\|f - g\|_b \le c_b$, and $\|f - g\|_{(\delta);a_{\mathcal{W}}} \le c_\delta$.*

Then, $h_{V^n}(f)$ exists, is finite, and satisfies the bound

$$|h_{V^n}(f) - h_{V^n}(g)| \le e^3\left[(2 + N\delta^{-1}) \vee \log\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right) \vee \log c_\delta^{-1} \vee \log c_b^{-1}\right] c_\delta$$
$$+2e^2(1-b^{-1})^{-1}(\|f+g\|_1 \vee 1)c_b$$

Proof. Combining Theorem 3.2.3 and Lemma 3.4.2 gives, for any α satisfying $\frac{N+\delta}{N+2\delta} \le \alpha < 1$, (and using $\alpha^{-1} < 2$):

$$\begin{aligned}
|h_{V^n}(f) - h_{V^n}(g)| &\le e\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right)^{1-\alpha}(1-\alpha)^{-1}\|f-g\|_{(\delta);a_{\mathcal{W}}}^{\alpha} \\
&\quad + 2e(1-b^{-1})^{-1}\|f+g\|_1^{1-\alpha b^{-1}}\|f-g\|_b^{\alpha} \\
&\le e\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right)^{1-\alpha}(1-\alpha)^{-1}c_\delta^{\alpha} + 2e(1-b^{-1})^{-1}(\|f+g\|_1 \vee 1)c_b^{\alpha}
\end{aligned}$$

Take $\alpha := 1 - \left[(2+N\delta^{-1}) \vee \log\left(\frac{\kappa_{N,n}\omega_n}{a_{\mathcal{W}}^n}\right) \vee \log c_\delta^{-1} \vee \log c_b^{-1}\right]^{-1} \ge \frac{N+\delta}{N+2\delta}$. □

Remark. Here is a typical application: Suppose P_ε is an ε-indexed sequence in $\hat{\mathcal{P}}(V^n)$ for $\varepsilon \in [0,1]$ satisfying an average energy constraint, and $h_{V^n}(P_0)$ exists and is finite. If $\|p_\varepsilon - p_0\|_b$ and $\|p_\varepsilon - p_0\|_{(\delta)}$ are $O(\varepsilon^l)$ as $\varepsilon \to 0$, then $h_{V^n}(P_\varepsilon) = h_{V^n}(P_0) + O(\varepsilon^l \log \frac{1}{\varepsilon})$.

Theorem 3.4.5. *Let $\mathcal{W}$ be semi-uniform with $a_{\mathcal{W}} \ge c_{\mathcal{W}}^o$. Let $\delta > 0$, $b \in (1,\infty]$. Suppose $f, g \in L_+^1(V^n)$, $h_{V^n}(g)$ exists and is finite, $\|f-g\|_b \le c_b < \infty$, and $\|f-g\|_{(\delta);a_{\mathcal{W}}} \le c_\delta < 1$.*

Define the quantities $c_\pm := \left|\int f \pm g\, d\mu\right|$, $b' := \frac{1}{1-b^{-1}}$, $K := \log^+\left(\kappa_{N,n}\omega_n c_b^{b'} a_{\mathcal{W}}{}^{-n}\right)/\log c_\delta^{-1}$, $\alpha_0 := K/(1+K) \in [0,1)$, and the function $\beta(\alpha) := \frac{\alpha(1-b^{-1})}{1-\alpha b^{-1}}[\alpha - (1-\alpha)K]$.

Then $h_{V^n}(f)$ exists and is finite, and we have the bound

$$|h_{V^n}(f) - h_{V^n}(g)| \le (c_+ \vee 1)e^2(K+1+2b')\left(\frac{1}{1-\beta\left(\frac{N+\delta}{N+2\delta}\right)} \vee \log c_\delta^{-1}\right)c_\delta$$
$$+e^3\left[1 + \left|\log \frac{c_+ a_{\mathcal{W}}^n}{\kappa_{N,n}\omega_n}\right| + \left(K+b'\right)\log c_\delta^{-1}\right]c_\delta^{1+\frac{\delta}{N+\delta}}$$

In the special case of $\|f\|_1 = \|g\|_1 = 1$ (when f, g are probability densities), we have the simpler bound

$$|h_{V^n}(f) - h_{V^n}(g)| \le 2e^2(K+1+2b')\left(\frac{1}{1-\beta\left(\frac{N+\delta}{N+2\delta}\right)} \vee \log c_\delta^{-1}\right)c_\delta$$

Proof. Without loss of generality we assume $c_b \ge \left(\kappa_{N,n}\omega_n a_{\mathcal{W}}{}^{-n}\right)^{b^{-1}-1}$. $\beta(\alpha)$ was defined so that

$$\left(\kappa_{N,n}\omega_n a_{\mathcal{W}}{}^{-n} c_b^{b'}\right)^{\frac{\alpha(1-\alpha)(1-b^{-1})}{1-\alpha b^{-1}}} c_\delta^{\frac{\alpha^2(1-b^{-1})}{1-\alpha b^{-1}}} = e^{\beta(\alpha)\log c_\delta} = c_\delta^{\beta(\alpha)}$$

Combining Theorem 3.2.4 and Lemma 3.4.2 we have, for all $\frac{N+\delta}{N+2\delta} \le \alpha < 1$

$$|h_{V^n}(f) - h_{V^n}(g)| \le \frac{c_+^{1-\alpha} e^{\frac{\alpha(2-\alpha)(1-b^{-1})}{1-\alpha b^{-1}}}(1-\alpha b^{-1})}{\alpha(1-\alpha)(1-b^{-1})} c_\delta^{\beta(\alpha)} +$$
$$+\left|\log c_+ - (1-\alpha) + \frac{1}{1-\alpha b^{-1}}\log\frac{a_{\mathcal{W}}^{n(1-\alpha)} c_b^\alpha}{(\kappa_{N,n}\omega_n)^{1-\alpha} c_\delta^\alpha}\right| c_-$$

It is easy to verify that $\beta(\alpha_0) = 0$, $\beta(1) = 1$, and β is increasing on $[\alpha_0, 1]$, thus $\beta^{-1}\colon [0,1] \to [\alpha_0, 1]$ is well-defined. Take

$$\alpha = \beta^{-1}\left(1 - \left[\left(1-\beta\left(\tfrac{N+\delta}{N+2\delta} \vee \alpha_0\right)\right)^{-1} \vee \log c_\delta^{-1}\right]^{-1}\right) \in [\alpha_0 \vee \tfrac{N+\delta}{N+2\delta}, 1)$$

which gives $\beta(\alpha)\log c_\delta = \log c_\delta + \frac{\log c_\delta^{-1}}{\left(1-\beta\left(\frac{N+\delta}{N+2\delta}\vee\alpha_0\right)\right)^{-1}\vee\log c_\delta^{-1}} \le \log c_\delta + 1$, so in the above bound we have $\left(\kappa_{N,n}\omega_n a_{\mathcal{W}}{}^{-n} c_b^{b'}\right)^{\frac{\alpha(1-\alpha)(1-b^{-1})}{1-\alpha b^{-1}}} c_\delta^{\frac{\alpha^2(1-b^{-1})}{1-\alpha b^{-1}}} \le e\, c_\delta$.

The remaining parts of the bound are simplified as follows: $\frac{\alpha(2-\alpha)(1-b^{-1})}{1-\alpha b^{-1}} \le \alpha(2-\alpha) \le 1$; $c_+^{1-\alpha} \le c_+ \vee 1$; With some algebra, and using $\alpha \ge \frac{N+\delta}{N+2\delta} \ge \frac{1}{2}$,

$$\begin{aligned}\frac{1-\alpha b^{-1}}{\alpha(1-\alpha)(1-b^{-1})} &= \frac{K+1+b'/\alpha}{1-\beta}\\ &\le (K+1+2b')\left(\frac{1}{1-\beta\left(\frac{N+\delta}{N+2\delta}\vee\alpha_0\right)}\vee\log c_\delta^{-1}\right)\end{aligned}$$

With some additional algebra and using $\frac{1}{1-\alpha b^{-1}} \le b'$, we have

$$\begin{aligned}|h_{V^n}(f)-h_{V^n}(g)| \le{}& (c_+\vee 1)e^2(K+1+2b')\left(\frac{1}{1-\beta\left(\frac{N+\delta}{N+2\delta}\right)}\vee\log c_\delta^{-1}\right)c_\delta +\\ &+\left[1+\left|\log\frac{c_+ a_{\mathcal{W}}^n}{\kappa_{N,n}\omega_n}\right| + \log^+\left(\kappa_{N,n}\omega_n a_{\mathcal{W}}^{-n} c_b^{b'}\right) + b'\log c_\delta^{-1}\right]c_-\end{aligned}$$

If $\|f\|_1 = \|g\|_1 = 1$ then $c_+ = 2$ and $c_- = 0$. Otherwise, by Hölder's inequality applied to $|f-g| = |f-g|^{\frac{\alpha(1-b^{-1})}{1-\alpha b^{-1}}}|f-g|^{\frac{1-\alpha}{1-\alpha b^{-1}}}$, we have

$$\begin{aligned}c_- \le \|f-g\|_1 &\le c_\alpha^{\frac{1-b^{-1}}{1-\alpha b^{-1}}} c_b^{\frac{1-\alpha}{1-\alpha b^{-1}}} \le \left[e\, c_\delta^\alpha\left(\kappa_{N,n}\omega_n a_{\mathcal{W}}^{-n} c_b^{b'}\right)^{1-\alpha}\right]^{\frac{1-b^{-1}}{1-\alpha b^{-1}}}\\ &\le e^{\frac{1-b^{-1}}{1-\alpha b^{-1}}} c_\delta^{\beta/\alpha} \le e^{\frac{1-b^{-1}}{1-\alpha b^{-1}}}(e\, c_\delta)^{1/\alpha} \le e^3 c_\delta^{1+\frac{\delta}{N+\delta}}\end{aligned}$$

which allows us to express the bound in terms of c_δ. $\square$

Corollary 3.4.6. *Let $\mathcal{W}$ be semi-uniform with $a_{\mathcal{W}} \ge c_{\mathcal{W}}^o$. Let $\delta > 0$, $b \in (1,\infty]$. Suppose $p \in \hat{\mathcal{P}}(V^n)$ and $g \in L_+^1(V^n)$, $g \not\equiv 0$, and define the probability density $q \in \hat{\mathcal{P}}(V^n)$ by normalizing g: $q = \|g\|_1^{-1} g$. Suppose $\|p-g\|_b \le c_b < \infty$ and $\|p-g\|_{(\delta);a_{\mathcal{W}}} \le$*

$c_\delta < \infty$.

If we also have (a) $\|p\|_b < C_b < \infty$ *and* $\|p\|_{(\delta);a_{\mathcal{W}}} < C_\delta < \infty$*, or (b) If* $\|g\|_b < C_b < \infty$ *and* $\|g\|_{(\delta);a_{\mathcal{W}}} < C_\delta < \infty$*, then* $h_{V^n}(p)$ *and* $h_{V^n}(q)$ *exist and are finite.*

If either (a) or (b) holds, define the quantities $c_- := |1 - \|g\|_1|$, $k_b := \Big(\|g\|_1^{-1} \vee 1\Big)c_b + \Big(\|g\|_1^{-1} C_b\Big)c_-$, $k_\delta := \Big(\|g\|_1^{-1} \vee 1\Big)c_\delta + \Big(\|g\|_1^{-1} C_\delta\Big)c_-$.

If $k_\delta < 1$*, also define* $b' := \frac{1}{1-b^{-1}}$, $K := \log^+\Big(\kappa_{N,n}\omega_n k_b^{b'} a_{\mathcal{W}}{}^{-n}\Big)/\log k_\delta^{-1}$, $\alpha_0 := K/(1+K) \in [0,1)$*, and the function* $\beta(\alpha) := \frac{\alpha(1-b^{-1})}{1-\alpha b^{-1}}[\alpha - (1-\alpha)K]$.

We then have the bound

$$|h_{V^n}(p) - h_{V^n}(q)| \leq 2e^2(K + 1 + 2b')\left(\frac{1}{1 - \beta\Big(\frac{N+\delta}{N+2\delta}\Big)} \vee \log k_\delta^{-1}\right) k_\delta$$

Proof. This follows from the previous theorem applied to p and q, with the observation that $p - q = (p - g) + (1 - \|g\|_1)\|g\|_1^{-1} g = \|g\|_1^{-1}(p - g) + (1 - \|g\|_1^{-1})p$, and the triangle inequality to get $\|p - q\|_b \leq k_b$ and $\|p - q\|_{(\delta);a_{\mathcal{W}}} \leq k_\delta$. $\square$

We end with a key theorem for localizing our analysis in the next chapter.

Theorem 3.4.7 (Cutoff Theorem)**.** *Let* $\mathcal{W}$ *be semi-uniform of dimension* $n \geq 1$, $a_{\mathcal{W}} \geq c^o_{\mathcal{W}}$, $R > 0$*, and* $\delta \in (0,1]$*. Let* $\psi \in L^1_+[R,\infty)$ *be differentiable with* $\psi' \in L^1[R,\infty)$, $\psi'(r) \leq -\delta\psi(r)/r$ *for all* $r \geq R$*, and* $\int_R^\infty r^{N-1+\delta}\psi(r)\,dr < \infty$.

(i) For all $\gamma \in [0,\delta]$*, define the constants*

$$\mathfrak{b}_\gamma := \sup_{y\in\mathbb{R}^N} \int_{|w-y|>R} |w-y|^\gamma \psi(|w-y|)\,dV^n(w)$$

We have the bound:

$$\mathfrak{b}_\gamma \leq \kappa_{N,n}\psi(R)\omega_n R^{n+\gamma}(1+c^o_{\mathcal{W}}R)^{n'} + \\ + \kappa_{N,n}N\omega_n \int_R^\infty \psi(r)r^{n-1+\gamma}(1+c^o_{\mathcal{W}}r)^{n'}\,dr$$

(ii) Let $\mu \in \hat{\mathcal{P}}(\mathbb{R}^N)$ with $\|\mu\|_{(\delta);a_{\mathcal{W}}} < \infty$. Suppose $f, g \in L^1_+(V^n)$, $\|g\|_\infty < \infty$, $\|g\|_{(\delta);a_{\mathcal{W}}} < \infty$ and suppose that for all $w \in \mathcal{W}$,

$$|f(w) - g(w)| \leq \int_{|y-w|>R} \psi(|y-w|)\,d\mu(y)$$

Then $h_{V^n}(f)$ and $h_{V^n}(g)$ exist and are finite, and

$$\begin{aligned} \|f-g\|_\infty &\leq \psi(R) \\ \|f-g\|_1 &\leq \mathfrak{b}_0 \\ \|f-g\|_{(\delta);a_{\mathcal{W}}} &\leq \Big(a^\delta_{\mathcal{W}}\mathfrak{b}_\delta + \|\mu\|_{(\delta);a_{\mathcal{W}}}\mathfrak{b}_0\Big) \end{aligned}$$

(iii) In particular, let $\psi(r) = \varphi_{n,\varepsilon}(r) := (2\pi\varepsilon^2)^{-n/2}\exp\big(-r^2/2\varepsilon^2\big)$. For all $R \geq \sqrt{N+1}\,\varepsilon$,

$$\begin{aligned} \|f-g\|_\infty &\leq \left[\frac{1}{2\pi\varepsilon^2}\exp\left(-\frac{R^2}{n\varepsilon^2}\right)\right]^{n/2} \\ \|f-g\|_1 &\leq 4Nn^{-1/2}\left[\left(\frac{eN}{n'}\right)^{\frac{1}{2}}(1+c^o_{\mathcal{W}}R)\right]^{n'}\left[\frac{R^2}{n\varepsilon^2}\exp\left(-\frac{R^2}{n\varepsilon^2}\right)\right]^{n/2} \\ \|f-g\|_{(\delta)} &\leq 4Nn^{-\frac{1}{2}}\Big[\|\mu\|_{(\delta);a_{\mathcal{W}}} + 1 + a_{\mathcal{W}}R\Big]\left[\left(\frac{eN}{n'}\right)^{\frac{1}{2}}(1+c^o_{\mathcal{W}}R)\right]^{n'}\left[\frac{R^2}{n\varepsilon^2}\exp\left(-\frac{R^2}{n\varepsilon^2}\right)\right]^{n/2} \end{aligned}$$

Example 3.4.1. The bounds of (iii) are $O(e^{-R^2/C\varepsilon^2})$, so by Theorem 3.4.4, $|h_{V^n}(f) - h_{V^n}(g)|$

is too. For example, suppose X is a random variable on $\mathbb{R}^N$, $Z \sim \mathcal{N}(0, \varepsilon^2 I_N)$, and $Z' \sim P_{Z\,|\,(|Z|<R)}$. Setting $Y = W + Z$ and $Y' = W + Z'$, (iii) and Lemma 3.3.1 show that $h(Y) - h(Y') = O(e^{-R^2/C\varepsilon^2})$ (assuming $h(Y)$ is finite). Note that this is a significantly stronger statement than the straightforward observation $\mathbb{P}(|Z| \geq R) = O(e^{-R^2/C\varepsilon^2})$.

Proof. (i) Note that $\psi'(r) \leq -\gamma\psi(r)/r \Leftrightarrow \frac{d}{dr}[\psi(r)r^\gamma] \leq 0$, so, applying Lemma 3.4.1,

$$\int_{|w-y|>R} |w-y|^\gamma \psi(|w-y|)\, dV^n(w) \leq \kappa_{N,n}\omega_n \int_R^\infty -\frac{d}{dr}[\psi(r)r^\gamma] r^n (1 + c^o_{\mathcal{W}} r)^{n'}\, dr$$

Integrate by parts, ignoring the term at $r = \infty$ since $\int_R^\infty r^{N+\delta}\psi(r)\, dr < \infty \implies \lim_{r\to\infty} r^{N+\gamma}\psi(r) = 0$:

$$\begin{aligned}\mathfrak{b}_\gamma \leq{}& \kappa_{N,n}\psi(R)\omega_n R^{n+\gamma}(1 + c^o_{\mathcal{W}} R)^{n'} + \\ &+ \kappa_{N,n} \int_R^\infty \psi(r)\omega_n r^{n-1} r^\gamma (n + N c^o_{\mathcal{W}} r)(1 + c^o_{\mathcal{W}} r)^{n'-1}\, dr\end{aligned}$$

which is less than the stated bound.

(ii) The L^∞ bound is immediate. The L^1 bound is:

$$\begin{aligned}\int_{\mathcal{W}} |f(w) - g(w)|\, dV^n &\leq \int_{\mathcal{W}} \int_{|y-w|>R} \psi(|y-w|)\, d\mu(y)\, dV^n(w) \\ &\leq \int \left[\int_{|w-y|>R} \psi(|w-y|)\, dV^n(w) \right] d\mu(y) \leq \mathfrak{b}_0\end{aligned}$$

For the decay norm estimate, we have

$$
\begin{aligned}
\|f-g\|_{(\delta);a_{\mathcal{W}}} &\le \int\int_{|y-w|>R} (1+a_{\mathcal{W}}|w|)^{\delta}\psi(|y-w|)\,dV^n(w)\,d\mu(y) \\
&\le \int\int_{|y-w|>R} \Big[a_{\mathcal{W}}^{\delta}|y-w|^{\delta} + (1+a_{\mathcal{W}}|y|)^{\delta}\Big]\psi(|y-w|)\,dV^n(w)\,d\mu(y) \\
&\le \Big(a_{\mathcal{W}}^{\delta}\mathfrak{b}_{\delta} + \|\mu\|_{(\delta);a_{\mathcal{W}}}\mathfrak{b}_0\Big)
\end{aligned}
$$

The entropy is finite by Corollary 3.2.5.

(iii) Define random variables $Z^{(k)} \sim \mathcal{N}(0, I_k)$ for $k \in \{n, n+1, \ldots, N+1\}$. The Gaussian tail expectations can be written as

$$
\begin{aligned}
\int_R^{\infty} r^{k-1}\varphi_{n,\varepsilon}(r)\,dr &= \frac{(2\pi\varepsilon^2)^{\frac{k-n}{2}}}{k\omega_k}\int_R^{\infty} k\omega_k r^{k-1}\varphi_{\varepsilon}^{k}(r)\,dr \\
&= \frac{(2\pi\varepsilon^2)^{\frac{k-n}{2}}}{k\omega_k}\mathbb{P}\Big[\Big|Z^{(k)}\Big| \ge R\varepsilon^{-1}\Big]
\end{aligned}
$$

The Chernoff tail bound for $Z^{(k)}$ is $\mathbb{P}[\big|Z^{(k)}\big|^2 \ge kt] \le (te^{1-t})^{k/2}$ for $t \ge 1$, so when $R \ge \sqrt{k}\,\varepsilon$ we have

$$
\begin{aligned}
\int_R^{\infty} r^{k-1}\varphi_{n,\varepsilon}(r)\,dr &\le \frac{(2\pi\varepsilon^2)^{\frac{k-n}{2}}}{k\omega_k}\left[\frac{eR^2}{k\varepsilon^2}\exp\!\left(-\frac{R^2}{k\varepsilon^2}\right)\right]^{k/2} \\
&\le \frac{1}{k\omega_k}\left(\frac{2\pi e}{k}\right)^{k/2} R^k\varphi_{n,\varepsilon}(R)
\end{aligned}
$$

Using the Sterling bound for the Gamma function, extended to $[\frac{3}{2},\infty)$, we have $\omega_k^{-1} = \Gamma(1+k/2)\pi^{-k/2} \le \sqrt{e\pi k/2}\big(\frac{k}{2\pi e}\big)^{k/2} \implies \int_R^{\infty} r^{k-1}\varphi_{n,\varepsilon}(r)\,dr < \big(\frac{e\pi}{2n}\big)^{1/2}R^k\varphi_{n,\varepsilon}(R)$.

This estimate extends linearly to tail expectations of polynomials in r, giving the bound, for $\gamma \in [0,1]$ (using $r^\gamma \leq 1 + r$):

$$\begin{aligned}\int_R^\infty \varphi_{n,\varepsilon}(r) r^{n-1+\gamma}(1 + c^o_{\mathcal{W}} r)^{n'}\, dr &\leq \int_R^\infty \varphi_{n,\varepsilon}(r) r^{n-1}(1+r)(1 + c^o_{\mathcal{W}} r)^{n'}\, dr \\ &\leq \left(\frac{e\pi}{2n}\right)^{1/2} R^n(1+R)(1 + c^o_{\mathcal{W}} R)^{n'} \varphi_{n,\varepsilon}(R)\end{aligned}$$

So for $\gamma \in [0,\delta]$ and $R\varepsilon^{-1} \geq \sqrt{N+1}$,

$$\begin{aligned}\mathfrak{b}_\gamma &\leq \left[1 + N\left(\frac{e\pi}{2n}\right)^{\frac{1}{2}}\right] \kappa_{N,n}\omega_n R^n (1+R)(1 + c^o_{\mathcal{W}} R)^{n'} \varphi_{n,\varepsilon}(R) \\ &\leq \left[1 + \left(\frac{e\pi}{2}\right)^{\frac{1}{2}}\right]\left(\frac{e}{2}\right)^{\frac{1}{2}} N n^{-\frac{1}{2}} \left(\frac{N}{n}\right)^{\frac{n}{2}} \left(\frac{N}{n'}\right)^{\frac{n'}{2}} \left(\frac{2\pi e}{n}\right)^{\frac{n}{2}} R^n(1+R)(1 + c^o_{\mathcal{W}} R)^{n'} \varphi_{n,\varepsilon}(R) \\ &\leq 4Nn^{-\frac{1}{2}}(1+R)\left[\left(\frac{eN}{n'}\right)^{\frac{1}{2}}(1 + c^o_{\mathcal{W}} R)\right]^{n'} \left[\frac{R^2}{n\varepsilon^2}\exp\left(-\frac{R^2}{n\varepsilon^2}\right)\right]^{\frac{n}{2}}\end{aligned}$$

$\square$

4

Asymptotic Capacity Results

The primary objective of this chapter is to prove Theorem 4.3.1, which states that an AWGN channel with an average power constraint and scale-invariant alphabet constraint $X/|X| \in \Omega$ (Ω a smooth, compact $(n-1)$-dimensional submanifold of S^{n-1}, possibly with boundary), has high-SNR capacity

$$\mathsf{Cap}(\mathsf{SNR}) \approx \frac{n}{2}\log(1+\mathsf{SNR}) + \log\frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})}$$

A closely related result, applicable only to the special case of Grassmann manifolds, was proven in [14] in the context of multiple antenna channels, and includes an additional term corresponding to a noncoherent fading block channel model that we do not consider. The geometric "sphere packing" interpretation presented by Zheng an Tse applies equally well to our general result.

Let P_X be any probability measure for an AWGN channel $\mathcal{X} \to \mathcal{Y} = \mathbb{R}^N$ with average noise ε^2 per degree of freedom. Denoting the noise pdf by $p_{Z_\varepsilon}(z) = \varphi_\varepsilon^N(|z|) \equiv (2\pi\varepsilon^2)^{-\frac{N}{2}} e^{-|z|^2/2\varepsilon^2}$ we then have

$$p_{Y_\varepsilon}(y) = \int_{\mathcal{X}} \varphi_\varepsilon^N(|y-x|)\, dP_X(x) = \int_{\mathcal{X}} p_X(x)\varphi_\varepsilon^N(|y-x|)\, dV^n(x)$$

where the second equality holds whenever $p_X := \frac{dP_X}{dV^n}$ exists. Beginning our capacity calculation in the standard way for AWGN channels,

$$I(X;Y_\varepsilon) = h(Y_\varepsilon) - h(Y_\varepsilon|X) = h(Y_\varepsilon) - h(Z_\varepsilon) = h(Y_\varepsilon) - \frac{N}{2}\log\bigl(2\pi e\varepsilon^2\bigr) \tag{4.0.1}$$

Thus capacity will be achieved by the P_X which maximizes the corresponding $h(Y_\varepsilon)$ (subject to any code-level constraints imposed on P_X, such as an average power constraint). Heuristically, when noise corruption ε^2 is small, one expects $h(Y_\varepsilon)$ to be maximized by a P_X of maximal, or nearly-maximal entropy. This intuition is largely correct: we will show that, to a zeroth-order approximation in ε^2, maximizing $h_{V^n}(X)$ maximizes $h(Y_\varepsilon)$. For a more precise capacity approximation in ε^2, the geometry of the embedding $\mathcal{X} \subset \mathbb{R}^N$ also plays a role, and the optimal P_X may be a perturbation from the $h_{V^n}(X)$-maximizing distribution.

Even in the zeroth-order case, we require some mild geometric prerequisites to justify our conclusions, and the situation must be analyzed carefully. When $n < N$, P_X and P_{Y_ε} are supported on spaces of different dimensionality, and the entropies $h_{V^n}(X)$ and $h(Y_\varepsilon)$ are taken with respect to different measures. In fact, for a manifold $\mathcal{X}$ which fails to be uniform in the sense of the previous chapter, the P_X maximizing $h(Y_\varepsilon)$ need not be even approximately maximal in $h_{V^n}(X)$ for any $\varepsilon > 0$.

4.1 Preliminaries

In the next two sections $\mathcal{X} \subset \mathbb{R}^N$ will be assumed a smooth submanifold that is uniform in the sense of Definition 3.4.1 for some constant $c_{\mathcal{X}} < \infty$. This assumption holds automatically for compact $\mathcal{X}$, as well as many non-compact submanifolds. We will subsequently be able to extend the entropy and capacity estimates of this section to wider classes of $\mathcal{X} \subset \mathbb{R}^N$.

4.1.1 Entropy in a Tubular Parameterization

We will need the following lemma multiple times in the proceeding sections:

Lemma 4.1.1. *Let $x_y, x \in \mathcal{X}$ with $r = d_{\mathcal{X}}(x, x_y) < \rho_\top(\mathcal{X})$ and $\nu_y \in T^\perp_{x_y}\mathcal{X}$ with $|\nu_y| < \rho_\perp(\mathcal{X})$. For $y := x_y + \nu_y$ the euclidean distance can be written*

$$|y - x|^2 = |\nu_y|^2 + r^2[1 + \delta_\top(x) + \nu_y \cdot \delta_\perp(x)]$$

where $\delta_\top(x) \in \mathbb{R}$ and $\delta_\perp(x) \in T^\perp_{x_y}\mathcal{X}$. These quantities satisfy the bounds

$$|\delta_\top(x)| \leq \frac{1}{2}c_{\mathrm{II}}^2 r^2 \quad \text{and} \quad |\delta_\perp(x)| \leq c_{\mathrm{II}}. \tag{4.1.1}$$

Combined, we have the simplified bounds, valid for $r \leq \left(\sqrt{2}c_{\mathcal{X}}\right)^{-1}$:

$$\frac{1}{2}\left(|\nu_y|^2 + r^2\right) \leq |y - x|^2 \leq \frac{3}{2}\left(|\nu_y|^2 + r^2\right) \tag{4.1.2}$$

Proof. Let $x = x(r)$ be an arc-length parameterized geodesic with $x(0) = x_y$, and expand the function $f(r) := |\nu_y + x_y - x(r)|^2$ in a Taylor series about $r = 0$ up to a 2nd order error term. We have $f(r) = |\nu_y|^2 + [x(r) - x_y] \cdot [x(r) - x_y] - 2\nu_y \cdot [x(r) - x_y]$, and $f(0) = |\nu_y|^2$. Differentiating, $f'(r) = 2x'(r) \cdot [x(r) - x_y] - 2\nu_y \cdot x'(r)$, and $f'(0) = 0$. Using $|x'(r)| \equiv 1$ and Lemma 2.4.2(i),

$$\begin{aligned}\frac{1}{2}f''(r) &= \left|x'(r)\right|^2 + x''(r) \cdot [x(r) - x_y - \nu_y] \\ &= 1 + \mathrm{I\!I}_{x(r)}(x'(r), x'(r)) \cdot [x(r) - x_y] - \mathrm{I\!I}_{x(r)}(x'(r), x'(r)) \cdot \nu_y\end{aligned}$$

and the mean-value form of the Taylor remainder gives, for some $0 \leq s \leq r$,

$$f(r) = |\nu_y|^2 + r^2\big[1 + r^2\, \mathrm{I\!I}_{x(s)}(x'(s), x'(s)) \cdot [x(s) - x_y] - \mathrm{I\!I}_{x(s)}(x'(s), x'(s)) \cdot \nu_y\big]$$

Setting $\delta_\perp(x(r)) := -\operatorname{Proj}_{T^\perp_{x_y}\mathcal{X}}[\mathrm{I\!I}_{x(s)}(x'(s), x'(s))]$, its bound is immediate.

For $\delta_\top(x(r)) := \mathrm{I\!I}_{x(s)}(x'(s), x'(s)) \cdot [x(s) - x_y]$, Lemma 2.4.2(ii) gives, for some $t \in [0, s]$,

$$\delta_\top(x(r)) = \frac{1}{2}s^2\, \mathrm{I\!I}_{x(s)}(x'(s), x'(s)) \cdot \mathrm{I\!I}_{x(t)}(x'(t), x'(t))$$

which satisfies the stated bound.

For the final bound note that, by Young's inequality, $|\nu_y \cdot \delta_\perp(x)| \leq \frac{1}{2}r^{-2}|\nu_y|^2 + \frac{1}{2}r^2|\delta_\perp(x)|^2$, apply the previous estimates, and use $c_{\mathrm{I\!I}}^2 r^2 \leq \frac{1}{2}$. $\square$

4.1.2 Applying the Cutoff Theorem

Let $\mathcal{U}_R^{\mathcal{X}}$ be a tubular neighborhood of some radius $R \leq \left(\sqrt{2}\, c_{\mathcal{X}}\right)^{-1}$ about $\mathcal{X}$. Every $y \in \mathcal{U}_R^{\mathcal{X}}$ may be uniquely represented as $y = x_y + \nu_y$ with $x_y \in \mathcal{X}$, $\nu_y \in T^\perp_{x_y}\mathcal{X}$. In order to estimate $h_{m^N}(Y_\varepsilon)$ in this geometrically-attuned tubular parameterization, we replace the true probability density $p_{Y_\varepsilon}(y)$ with

$$f_{Y_\varepsilon}(y) := \begin{cases} \int_{B_R^{\mathcal{X}}(x_y)} \varphi_{N,\varepsilon}(|\nu_y + x_y - \tilde{x}|)\, dP_X(\tilde{x}), & y \in \mathcal{U}_R^{\mathcal{X}} \\ 0, & \text{otherwise.} \end{cases}$$

In order to apply the cutoff theorem, we prove:

Lemma 4.1.2. *Let* $R \leq \left(\sqrt{2}\, c_{\mathcal{X}}\right)^{-1}$ *and* $\delta \in (0,1]$. *If* $|\nu_y| \vee d_{\mathcal{X}}(x, x_y) \geq R$ *then* $|\nu_y + x_y - x| \geq \frac{R}{\sqrt{2}}$.

Proof. Let $\gamma\colon [0,1] \to \mathbb{R}^N$ be the straight line with $\gamma(0) = x$ and $\gamma(1) = y := x_y + \nu_y$. Let $U := \{y' \in \mathbb{R}^N \colon d_{\mathcal{X}}(x, x_{y'}) \vee |\nu_{y'}| < R\}$. Since $x \in U$ and $y \notin U$, there is a $t^* \in (0,1]$ with $\gamma(t^*) \in \partial U$, so $d_{\mathcal{X}}(x, x_{\gamma(t^*)}) \vee |\nu_{\gamma(t^*)}| = R$. By (4.1.2), $\frac{1}{2}R^2 \leq |\gamma(t^*) - x|^2$. Observing that $|\gamma(t^*) - x| \leq |y - x|$ completes the proof. □

By the preceding lemma and the definition of f_{Y_ε}, we have

$$|p_{Y_\varepsilon}(y) - f_{Y_\varepsilon}(y)| \leq \int_{|x-y| \geq \frac{R}{\sqrt{2}}} \varphi_{N,\varepsilon}(|y - \tilde{x}|)\, dP_X(\tilde{x})$$

We will apply the cutoff theorem (Theorem 3.4.7(iii)) with $\mathcal{W} = \mathbb{R}^N$ and $\mu = P_X$, followed by Theorem 3.2.3, to estimate the error in our entropy estimates incurred by this restriction. Once $\varepsilon \leq R(N+1)^{-\frac{1}{2}}$, the error decays rapidly in ε, as $O\big(\exp\big(-R^2/2\varepsilon^2\big)\big)$. Therefore, we focus our analysis on estimating and maximizing $h_{m^N}(f_{Y_\varepsilon})$, for which a tubular parameterization is available.

4.1.3 Definition of f_{X_ε}

At each $x \in \mathcal{X}$, $\mathsf{expm}_x(\tau)$ maps $\tau \in T_x\mathcal{X}$ with $|\tau| \leq R$ into $B_R^{\mathcal{X}}(x)$. When x is fixed we also use the polar notation $\tau = r\hat{\omega}$ and $\mathsf{expm}_x(r\hat{\omega}) = \gamma_{\hat{\omega}}(r)$, where, by definition of geodesic coordinates, $\gamma_{\hat{\omega}}$ is the arc-length parameterized geodesic with $\gamma_{\hat{\omega}}(0) = x$ and $\gamma'_{\hat{\omega}}(0) = \hat{\omega}$. We write $dm^n(\tau)$ for the infinitesimal Euclidean n-volume on $T_x\mathcal{X} \approxeq \mathbb{R}^n$, which relates to dV^n via the Jacobian factor $J_x(\tau) \equiv \frac{dV^n(\mathsf{expm}_x)}{dm^n}(\tau)$.

In order to state our first main result we will need to define the auxiliary function $f_{X_\varepsilon} \in \mathcal{P}(V^n)$ as

$$f_{X_\varepsilon}(x) := \int_{B_R^{\mathcal{X}}(x)} \varphi_{n,\varepsilon}(d_{\mathcal{X}}(x,\tilde{x}))\, J_x(\tilde{x})^{-1}\, dP_X(\tilde{x})$$

Note that if $\mathcal{X}$ is a flat plane, and in the limit of $R \to \infty$, this definition reduces to the n-dimensional convolution of P_X with the $\mathcal{N}(0, \varepsilon^2 I_n)$ Gaussian distribution. In general f_{X_ε} may be considered to be a type of n-dimensional smoothing of P_X. The properties of f_{X_ε} required for our results are proven in Section 4.2.1 below.

4.2 $I(X; Y_\varepsilon)$ for General P_X; Capacity when $\mathcal{X}$ Compact

Theorem 4.2.1 (Asymptotic Mutual Information For General P_X)**.** *Let $\mathcal{X}$ be a smooth n-dimensional submanifold of $\mathbb{R}^N$, that is uniform in the sense of Definition 3.4.1, with $N \geq 1$ and $0 < c_{\mathcal{X}} < \infty$. Let $\delta \in (0, 1]$ and require $P_X \in \hat{\mathcal{P}}(\mathcal{X})$ with $\mathbb{E}|X|^\delta < \infty$. Suppose $Y_\varepsilon = X + Z$ where $Z \sim \mathcal{N}(0, \varepsilon^2 I_N)$ with $Z \perp X$, and assume $c_{\mathcal{X}}\varepsilon \leq (20 \vee 2(N+1))^{-1/2}$. Then,*

$$\left| I(X; Y_\varepsilon) - \left[\frac{n}{2} \log \frac{1}{2\pi e \varepsilon^2} + h_{V^n}(f_{X_\varepsilon})\right] \right| \leq \mathsf{const}(n, N)\delta^{-1} \|P_X\|_{(\delta)} (c_{\mathcal{X}}\varepsilon)^2 \log^2 (c_{\mathcal{X}}\varepsilon)^{-1}$$

Suppose, in addition, that $p_X := \frac{dP_X}{dV^n} \in \hat{\mathcal{P}}(V^n)$ *exists and is* $C^2(\mathcal{X})$*. Define the function* $D^2 p_X \in \mathcal{P}(V^n)$ *by*

$$D^2 p_X(x) := \sup_{\{\gamma(t):\, \gamma(0)=x, |\gamma'(0)|=1\}} \left|(p_X \circ \gamma)''(0)\right|$$

If $\left\|D^2 p_X\right\|_{(\delta)} < \infty$ *and* $\left\|D^2 p_X\right\|_\infty < \infty$*, then*

$$\left| h_N(p_{Y_\varepsilon}) - \frac{n'}{2}\log(2\pi e\varepsilon^2) - h_{V^n}(P_X) \right| \leq \mathsf{const}(n,N)\delta^{-1}\|P_X\|_{(\delta)}(c_{\mathcal{X}}\varepsilon)^2 \log^2(c_{\mathcal{X}}\varepsilon)^{-1} +$$

$$+\mathsf{const}(n)\delta^{-1}\left\|D^2 p_X\right\|_{(\delta)}\varepsilon^2\left[1 + \log^+\left(c_{\mathcal{X}}^{-n}\left\|D^2 p_X\right\|_\infty \varepsilon^2\right)\right]$$

Proof. The details of certain estimates needed in the proof are given in subsections 4.2.1 and 4.2.2 below, in order to focus here on the overall approach.

Take $R = \left[\frac{4e}{e-1}\log(c_{\mathcal{X}}\varepsilon)^{-1}\right]^{1/2}\varepsilon$. When $c_{\mathcal{X}}\varepsilon \leq (20 \vee 2(N+1))^{-1/2}$ this choice satisfies the following easily-verified inequalities: $R < (\sqrt{2}\,c_{\mathcal{X}})^{-1}$, $\frac{R}{\sqrt{2}} > \sqrt{N+1}\,\varepsilon$, and (needed in evaluating cutoff theorem bounds)

$$\left[\frac{R^2}{n\varepsilon^2}\exp\left(-\frac{R^2}{n\varepsilon^2}\right)\right]^{n/2} \leq \exp\left(-\frac{R^2}{2\varepsilon^2}\left(1 - e^{-1}\right)\right) \leq (c_{\mathcal{X}}\varepsilon)^2$$

Define $\phi_{k,\varepsilon}^{(R)}(r) := 1_{[0,R]}(r)\phi_{k,\varepsilon}(r)$. Inside the tubular neighborhood $\mathcal{U}_R^{\mathcal{X}}$, using Lemma 4.1.1, and the shorthand $r := d_{\mathcal{X}}(x_y, \tilde{x})$, we have:

$$\begin{aligned}
f_{Y_\varepsilon}(y) &= \int_{B_R^{\mathcal{X}}(x_y)} \varphi_{N,\varepsilon}(|y - \tilde{x}|)\, dP_X(\tilde{x}) \\
&= \varphi_{n',\varepsilon}(|\nu_y|) \int_{B_R^{\mathcal{X}}(x_y)} \left(2\pi\varepsilon^2\right)^{-\frac{n}{2}} e^{-\frac{r^2}{2\varepsilon^2}[1+\delta_\top(\tilde{x})+\nu_y\cdot\delta_\perp(\tilde{x})]}\, dP_X(\tilde{x}) \\
&= \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi_{n,\varepsilon}^{(R)}(r) e^{-\frac{r^2}{2\varepsilon^2}\delta_\top} \cosh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right)\left[1 - \tanh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right)\right] dP_X(\tilde{x})
\end{aligned}$$

In Section 4.2.2 we define the function $g_{Y_\varepsilon} \in L^1_+(\mathcal{U}^{\mathcal{X}}_R)$, and its ν_y-even and odd parts, respectively:

$$
\begin{aligned}
g_{Y_\varepsilon}(y) &:= \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi^{(R)}_{n,\varepsilon}(r)\, J_x(\tilde{x})^{-1} \left[1 - \tanh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right)\right] dP_X(\tilde{x}) \\
g_e(y) &:= \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi^{(R)}_{n,\varepsilon}(r)\, J_x(\tilde{x})^{-1}\, dP_X(\tilde{x}) \equiv \varphi_{n',\varepsilon}(|\nu_y|)\, f_{X_\varepsilon}(x_y) \\
g_o(y) &:= \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi^{(R)}_{n,\varepsilon}(r)\, J_x(\tilde{x})^{-1} \tanh\left(-\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right) dP_X(\tilde{x})
\end{aligned}
$$

Lemma 4.2.8 of that section, the cutoff theorem, and the triangle inequality together give

$$
\begin{aligned}
\|p_{Y_\varepsilon} - g_{Y_\varepsilon}\|_\infty &\leq \mathsf{const}(n, N)\varepsilon^{-N} \\
\|p_{Y_\varepsilon} - g_{Y_\varepsilon}\|_1 &\leq \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2 \\
\|p_{Y_\varepsilon} - g_{Y_\varepsilon}\|_{(\delta);c_{\mathcal{X}}} &\leq \mathsf{const}(n, N)\delta^{-1}\|P_X\|_{(\delta)}(c_{\mathcal{X}}\varepsilon)^2
\end{aligned}
$$

Define the normalized probability density $q_{Y_\varepsilon} = \|g_{Y_\varepsilon}\|_1^{-1} g_{Y_\varepsilon}$. We may obtain an estimate of $h_N(p_{Y_\varepsilon}) - h_N(q_{Y_\varepsilon})$ by Theorem 3.2.3, and when $(c_{\mathcal{X}}\varepsilon)$ is sufficiently small, Corollary 3.4.6 applies, giving a bound

$$
|h_N(p_{Y_\varepsilon}) - h_N(q_{Y_\varepsilon})| \leq \mathsf{const}(n, N)\delta^{-1}\|P_X\|_{(\delta)}(c_{\mathcal{X}}\varepsilon)^2 \log(c_{\mathcal{X}}\varepsilon)^{-1}
$$

Similarly, Lemma 4.2.10 gives us

$$
\left| h_N(q_e) - \left[h_{V^n}(f_{X_\varepsilon}) + \frac{n'}{2}\log(2\pi e\varepsilon^2) \right] \right| \leq \mathsf{const}(n, N)\delta^{-1}\|P_X\|_{(\delta)}(c_{\mathcal{X}}\varepsilon)^2 \log(c_{\mathcal{X}}\varepsilon)^{-1}
$$

To complete the estimate of $h_N(p_{Y_\varepsilon})$ we need $|h_N(q_{Y_\varepsilon}) - h_N(q_e)|$, which is provided by

Lemma 4.2.11

$$|h_N(q_{Y_\varepsilon}) - h_N(q_e)| \leq \mathsf{const}(n,N)\delta^{-1}\|P_X\|_{(\delta)}c_{\mathcal{X}}{}^2\left[R^2+\varepsilon^2\right]\log(c_{\mathcal{X}}\varepsilon)^{-1}$$
$$\leq \mathsf{const}(n,N)\delta^{-1}\|P_X\|_{(\delta)}(c_{\mathcal{X}}\varepsilon)^2\log^2(c_{\mathcal{X}}\varepsilon)^{-1}$$

In the case when $p_X \in C^2(\mathcal{X})$ exists, the entropy estimate follows by Lemma 4.2.5 and Theorem 3.2.4, or Corollary 3.4.6 for sufficiently small ε. □

Theorem 4.2.2 (Asymptotic Channel Capacity for Compact Alphabets)**.** *Let $\mathcal{X}$ be a smooth, compact n-dimensional submanifold of $\mathbb{R}^N$ with $N \geq 1$ and diameter d. Define a communications channel $X \longrightarrow Y_\varepsilon = X + Z$ where $Z \sim \mathcal{N}(0, \varepsilon^2 I_N)$ and $Z \perp X$. If $c_{\mathcal{X}}\varepsilon \leq (20 \vee 2(N+1))^{-1/2}$ then the channel capacity (in nats) is approximated by*

$$\left|\mathsf{Cap}(\varepsilon) - \left[\frac{n}{2}\log\frac{1}{\varepsilon^2} + \log\frac{V^n(\mathcal{X})}{(2\pi e)^{n/2}}\right]\right| \leq \mathsf{const}(n,N)\big(1+\tfrac{1}{2}c_{\mathcal{X}}\mathsf{d}\big)(c_{\mathcal{X}}\varepsilon)^2\log^2(c_{\mathcal{X}}\varepsilon)^{-1}$$

Proof. Since $\mathcal{X}$ is compact, it is automatically uniform with $c_{\mathcal{X}} > 0$. For any $P_X \in \hat{\mathcal{P}}(\mathcal{X})$ we can apply Theorem 4.2.1 with $\delta = 1$. By shifting $\mathcal{X}$ by the appropriate vector in $\mathbb{R}^N$ we may assume $\|P_X\|_{(\delta)} \leq 1 + \frac{1}{2}c_{\mathcal{X}}\mathsf{d}$, and

$$\left|I(X;Y_\varepsilon) - \left[\frac{n}{2}\log\frac{1}{2\pi e\varepsilon^2} + h_{V^n}(f_{X_\varepsilon})\right]\right| \leq \mathsf{const}(n,N)\big(1+\tfrac{1}{2}c_{\mathcal{X}}\mathsf{d}\big)(c_{\mathcal{X}}\varepsilon)^2\log^2(c_{\mathcal{X}}\varepsilon)^{-1}$$

By Lemma 3.1.1 and Lemma 4.2.4,

$$\mathsf{Cap}(\varepsilon) - \left[\frac{n}{2}\log\frac{1}{\varepsilon^2} + \log\frac{V^n(\mathcal{X})}{(2\pi e)^{n/2}}\right] \leq \mathsf{const}(n,N)\big(1+\tfrac{1}{2}c_{\mathcal{X}}\mathsf{d}\big)(c_{\mathcal{X}}\varepsilon)^2\log^2(c_{\mathcal{X}}\varepsilon)^{-1}$$

To complete the proof we need only show that the claimed asymptotic capacity can be achieved (up to the stated error term). Take the volume-constant $P_X = V^n(\mathcal{X})^{-1}V^n$, giving $h_{V^n}(P_X) = \log V^n(\mathcal{X})$, and note that p_X is constant, hence $D^2 p_X = 0$. The bound now follows by the second half of Theorem 4.2.1. □

4.2.1 The f_{X_ε} Auxiliary Function

In the subsequent bounds we will use the following notation:

$$\xi(\rho) := \left[\frac{c_{\mathcal{X}}\rho}{\sin(c_{\mathcal{X}}\rho)}\right]^{n-1} \qquad \eta(\rho) := \left[\frac{\sinh(c_{\mathcal{X}}\rho)}{\sin(c_{\mathcal{X}}\rho)}\right]^{n-1}$$

$$\eta_{2k,n,c_{\mathcal{X}}\varepsilon} := \sum_{l=0}^{n-1}\binom{n-1}{l}\frac{(n+2k+2l-2)!!}{(n+2k-2)!!\,2^l}(c_{\mathcal{X}}\varepsilon)^{2l}$$

The functions $\xi(\rho)$ and $\eta(\rho)$ come into play using Lemma 2.4.3 and Theorem 2.4.4 to bound J_x terms. Below we apply these bounds as needed without further comment. $\eta_{k,n,c_{\mathcal{X}}\varepsilon}$ is a bounding constant of the form $\eta_{k,n,c_{\mathcal{X}}\varepsilon} = 1 + O\Big((c_{\mathcal{X}}\varepsilon)^2\Big)$; It will arise below from the following:

Lemma 4.2.3. *For $R \le \left(\sqrt{2}\,c_{\mathcal{X}}\right)^{-1}$,*

$$\int \eta(r)\left(\frac{r}{\varepsilon}\right)^{2k}\chi^{(R)}_{n,\varepsilon}(r)\,dr \le \frac{(n+2k-2)!!}{(n-2)!!}\eta_{2k,n,c_{\mathcal{X}}\varepsilon}$$

Proof. By Taylor's theorem, for $0 \le x \le 2^{-1/2}$, $\sinh x \le x + \frac{\cosh(2^{-1/2})}{6}x^2$, and $\sin x \ge 1 - \frac{1}{6}x^2$, so $\frac{\sinh x}{\sin x} \le \frac{1+\frac{1}{6}\cosh(2^{-1/2})x^2}{1-\frac{1}{6}x^2} \le 1 + \frac{1+\cosh(2^{-1/2})}{6-x^2}x^2 \le 1 + \frac{1}{2}x^2$ and $\eta(r) \le \left(1+\frac{1}{2}c_{\mathcal{X}}^2 r^2\right)^{n-1} = \sum_{l=0}^{n-1}\binom{n-1}{l}\left(\frac{c_{\mathcal{X}}r}{\sqrt{2}}\right)^{2l}$. Plugging this in to the integral and using $\int r^{2m}\chi_{n,\varepsilon}(r)\,dr = \frac{(n-2+2m)!!}{(n-2)!!}\varepsilon^{2m}$ gives the result. □

The following properties of f_{X_ε} will be needed in the main results:

Lemma 4.2.4. *For* $R \le \left(\sqrt{2}\, c_{\mathcal{X}}\right)^{-1}$ *and* $\delta \in (0,1]$ *we have*

$$|1 - \|f_{X_\varepsilon}\|_1| \le \eta_{0,n,c_{\mathcal{X}}\varepsilon} - 1 = \frac{n(n-1)}{2}(c_{\mathcal{X}}\varepsilon)^2 + O((c_{\mathcal{X}}\varepsilon)^4)$$
$$\|f_{X_\varepsilon}\|_{(\delta)} \le 2\eta_{0,n,c_{\mathcal{X}}\varepsilon}\|P_X\|_{(\delta)} \le 2\eta(2^{-\frac{1}{2}})\|P_X\|_{(\delta)}$$

Proof. Rearranging order of integration, changing to geodesic coordinates centered at $\tilde{x}$, and using the normalization properties of the integrals:

$$\begin{aligned}
\|f_{X_\varepsilon}\|_1 &= \iint \varphi_{n,\varepsilon}^{(R)}(d_{\mathcal{X}}(x,\tilde{x}))\, J_x(\tilde{x})^{-1}\, dP_X(\tilde{x})\, dV^n(x) \\
&= \int \left[\int \varphi_{n,\varepsilon}^{(R)}(d_{\mathcal{X}}(x,\tilde{x}))\, J_x(\tilde{x})^{-1}\, dV^n(x)\right] dP_X(\tilde{x}) \\
&= \int \left[\int \varphi_{n,\varepsilon}^{(R)}(|\tau|)\, \frac{J_{\tilde{x}}(\tau)}{J_{\mathsf{expm}_{\tilde{x}}\,\tau}(\tilde{x})}\, dm^n(\tau)\right] dP_X(\tilde{x}) \\
&= 1 + \int \left[\int \varphi_{n,\varepsilon}^{(R)}(|\tau|)\left(\frac{J_{\tilde{x}}(\tau)}{J_{\mathsf{expm}_{\tilde{x}}\,\tau}(\tilde{x})} - 1\right) dm^n(\tau)\right] dP_X(\tilde{x})
\end{aligned}$$

For any $a > 0$ we have $a - 1 \ge 1 - a^{-1}$, so changing to geodesic polar coordinates $\tau = r\hat{\omega}$ and taking $a = \eta(r)$, note that

$$\left|\frac{J_{\tilde{x}}(\tau)}{J_{\mathsf{expm}_{\tilde{x}}\,\tau}(\tilde{x})} - 1\right| \le [\eta(r) - 1] \vee \left[1 - \eta(r)^{-1}\right] \le \eta(r) - 1$$

Therefore, $|1 - \|f_{X_\varepsilon}\|_1| \le \int \chi_{n,\varepsilon}^{(R)}(r)\,[\eta(r) - 1]\, dr = \eta_{0,n,c_{\mathcal{X}}\varepsilon} - 1$.

Similarly for the decay norm, using $(1 + c_{\mathcal{X}}|x|)^\delta \le \left[(1 + c_{\mathcal{X}}|\tilde{x}|)^\delta + (c_{\mathcal{X}}|x - \tilde{x}|)^\delta\right]$,

$(c_{\mathcal{X}} r)^\delta \leq 2^{-\delta/2} < 1$, and $\|P_X\|_{(\delta)} \geq 1$:

$$\begin{aligned}
\|f_{X_\varepsilon}\|_{(\delta)} &= \iint (1 + c_{\mathcal{X}}|x|)^\delta \varphi^{(R)}_{n,\varepsilon}(d_{\mathcal{X}}(x,\tilde{x}))\, J_x(\tilde{x})^{-1}\, dP_X(\tilde{x})\, dV^n(x)\, dP_X(\tilde{x}) \\
&\leq \int (1 + c_{\mathcal{X}}|\tilde{x}|)^\delta \left[\int \chi^{(R)}_{n,\varepsilon}(r)\, \eta(r)\, dr\right] dP_X(\tilde{x}) + \\
&\quad + \iint (c_{\mathcal{X}}|r|)^\delta \left[\int \chi^{(R)}_{n,\varepsilon}(r)\, \eta(r)\, dr\right] dP_X(\tilde{x}) \\
&\leq \eta_{0,n,c_{\mathcal{X}}\varepsilon}\left[\|P_X\|_{(\delta)} + 1\right] \leq 2\eta_{0,n,c_{\mathcal{X}}\varepsilon}\|P_X\|_{(\delta)}
\end{aligned}$$

□

Lemma 4.2.5. *Suppose $p_X := \frac{dP_X}{dV^n}$ exists and is $C^2(\mathcal{X})$, and $\delta \in (0,1]$. We have:*

$$\begin{aligned}
\|f_{X_\varepsilon} - p_X\|_\infty &\leq \frac{n}{2}\left\|D^2 p_X\right\|_\infty \varepsilon^2 \\
\|f_{X_\varepsilon} - p_X\|_1 &\leq \eta_{2,n,c_{\mathcal{X}}\varepsilon}\, \frac{n}{2}\left\|D^2 p_X\right\|_1 \varepsilon^2 \\
\|p_{X_\varepsilon} - p_X\|_{(\delta)} &\leq \eta_{2,n,c_{\mathcal{X}}\varepsilon}\, n\left\|D^2 p_X\right\|_{(\delta)} \varepsilon^2
\end{aligned}$$

Proof. Fix x and change to a geodesic polar coordinate system centered on it, denoted $\tau = r\hat{\omega} \in T_x\mathcal{X}$:

$$\begin{aligned}
f_{X_\varepsilon}(x) &= \int \varphi^{(R)}_{n,\varepsilon}(d_{\mathcal{X}}(x,\tilde{x}))\, J_x(\tilde{x})^{-1}\, p_X(\tilde{x})\, dV^n(\tilde{x}) \\
&= \int \varphi^{(R)}_{n,\varepsilon}(|\tau|)\, p_X(\mathsf{expm}_x(\tau))\, dm^n(\tau) \\
&= \int \chi^{(R)}_{n,\varepsilon}(r)\left[\frac{1}{n\omega_n}\int_{S^{n-1}} p_X(\mathsf{expm}_x(r\hat{\omega}))\, d\hat{\omega}\right] dr
\end{aligned}$$

where $d\hat{\omega}$ is the standard $(n-1)$-dimensional measure on solid angles $\hat{\omega} \in S^{n-1} \subset \mathbb{R}^n$.

Using $\int \chi_{n,\varepsilon}^{(R)}(r)\,dr = 1$ and the fundamental theorem of calculus,

$$\begin{aligned}
f_{X_\varepsilon}(x) - p_X(x) &= \int \chi_{n,\varepsilon}^{(R)}(r)\left[\frac{1}{n\omega_n}\int_{S^{n-1}} p_X(\gamma_{\hat\omega}(r)) - p_X(x)\,d\hat\omega\right] dr \\
&= \int \chi_{n,\varepsilon}^{(R)}(r)\left[\frac{1}{n\omega_n}\int_{S^{n-1}}\int_0^r D_{\gamma'_{\hat\omega}(t)}[p_X(\gamma_{\hat\omega}(t))]\,dt\,d\hat\omega\right] dr \\
&= \int \chi_{n,\varepsilon}^{(R)}(r)\int_0^r\left[\frac{1}{n\omega_n}\int_{S^{n-1}} D_{\gamma'_{\hat\omega}(t)}[p_X(\gamma_{\hat\omega}(t))]\,d\hat\omega\right] dt\,dr
\end{aligned}$$

For brevity, we write the integral over r as a probability expectation, and take absolute value, yielding

$$|f_{X_\varepsilon}(x) - p_X(x)| \le \mathbb{E}_r\left[\int_0^r \frac{1}{n\omega_n}\left|\int_{S^{n-1}} D_{\gamma'_{\hat\omega}(t)}[p_X(\gamma_{\hat\omega}(t))]\,d\hat\omega\right| dt\right]$$

By symmetry, $\int_{S^{n-1}} D_{\gamma'_{\hat\omega}(0)}[p_X(\gamma_{\hat\omega}(0))]\,d\hat\omega = 0$, so we can use the FTC again:

$$\begin{aligned}
|f_{X_\varepsilon}(x) - p_X(x)| &\le \mathbb{E}_r\left[\int_0^r\int_0^t \frac{1}{n\omega_n}\left|\int_{S^{n-1}} D^2_{\gamma'_{\hat\omega}(u)\gamma'_{\hat\omega}(u)}[p_X(\gamma_{\hat\omega}(u))]\,d\hat\omega\,du\right| dt\right] \\
&\le \mathbb{E}_r\left[\int_0^r\int_0^t \frac{1}{n\omega_n u^{n-1}}\int_{S_u^{n-1}} \left|D^2 p_X(\gamma_{\hat\omega}(u))\right| d\sigma^{n-1}\,du\,dt\right] \qquad (4.2.1)
\end{aligned}$$

where σ^{n-1} is the standard $(n-1)$-dimensional volume measure on the euclidean sphere of radius u, and in the final line we have used the fact that $|\gamma'_{\hat\omega}| \equiv 1$. From (4.2.1) we easily get the L^∞ bound:

$$|f_{X_\varepsilon}(x) - p_X(x)| \le \mathbb{E}_r\left[\int_0^r\int_0^t \left\|D^2 p_X\right\|_\infty du\,dt\right] = \frac{n}{2}\left\|D^2 p_X\right\|_\infty \varepsilon^2$$

We now turn to estimating $\|f_{X_\varepsilon} - p_X\|_{(\delta)}$ with $\delta \in [0,1]$ (The L^1 bound will be given by the case $\delta = 0$.) Again for brevity, we will use the following temporary notation below ($\rho > 0$ and $x, x' \in \mathcal{X}$:

$$\psi(\rho, x, x') := \left|B^n_\rho\right|^{-1} 1_{[0,\rho]}(d_{\mathcal{X}}(x, x'))\left|D^2 p_X(x')\right|$$

Note that $\int_{S^{n-1}_u} f(u\hat{\omega})\, d\sigma^{n-1} = \frac{d}{du}\left[\int_{B^n_u} f(u\hat{\omega})\, dm^n\right]$, so, returning to the inner integrals of (4.2.1), we can apply integration by parts, followed by changing variables to integrate over $\mathcal{X}$:

$$\begin{aligned}
&\int_0^t \frac{1}{n\omega_n u^{n-1}} \int_{S^{n-1}_u} \left|D^2 p_X(\gamma_{\hat{\omega}}(u))\right| d\sigma^{n-1}\, du = \\
&\quad = \frac{1}{n\omega_n t^{n-1}} \int_{B^n_t} \left|D^2 p_X(\gamma_{\hat{\omega}}(t))\right| dm^n + \int_0^t \frac{n-1}{n\omega_n u^n} \int_{B^n_u} \left|D^2 p_X(\gamma_{\hat{\omega}}(u))\right| dm^n\, dt \\
&\quad = \frac{t}{n|B^n_t|} \int_{B^{\mathcal{X}}_t(x)} \frac{\left|D^2 p_X\right|}{J_x \circ \mathsf{expm}_x^{-1}}\, dV^n + \int_0^t \frac{n-1}{n|B^n_u|} \int_{B^{\mathcal{X}}_u(x)} \frac{\left|D^2 p_X\right|}{J_x \circ \mathsf{expm}_x^{-1}}\, dV^n\, du \\
&\quad \leq \frac{t\,\xi(t)}{n} \int_{\mathcal{X}} \psi(t, x, x')\, dV^n(x') + \frac{n-1}{n} \int_0^t \xi(u) \int_{\mathcal{X}} \psi(u, x, x')\, dV^n(x')\, du \\
&\quad \leq t\,\xi(t) \int_{\mathcal{X}} \psi(t, x, x')\, dV^n(x')
\end{aligned}$$

Integrating (4.2.1) and rearranging order of integration, we now have

$$\|f_{X_\varepsilon} - p_X\|_{(\delta)} \leq \mathbb{E}_r\left[\int_0^r t\,\xi(t) \iint (1 + c_{\mathcal{X}}|x|)^\delta \psi(t, x, x')\, dV^n(x')\, dV^n(x)\, dt\right] \tag{4.2.2}$$

Note that if we could replace $|x|$ with $|x'|$ in the above, we could use

$$\begin{aligned}
\iint (1 + c_{\mathcal{X}}\left|x'\right|)^\delta \psi(\rho, x, x')\, dV^n(x')\, dV^n(x) &= \iint (1 + c_{\mathcal{X}}\left|x'\right|)^\delta \psi(\rho, x, x')\, dV^n(x)\, dV^n(x') \\
&= \left|B^n_\rho\right|^{-1} \int (1 + c_{\mathcal{X}}\left|x'\right|)^\delta \left|D^2 p_X(x')\right| \int_{B^{\mathcal{X}}_\rho(x')} dV^n(x)\, dV^n(x') \\
&\leq \left[\frac{\sinh(c_{\mathcal{X}}\rho)}{c_{\mathcal{X}}\rho}\right]^{n-1} \left\|D^2 p_X\right\|_{(\delta)}
\end{aligned} \tag{4.2.3}$$

In the $\delta = 0$ case, $(1 + c_{\mathcal{X}}|x|)^\delta = (1 + c_{\mathcal{X}}|x'|)^\delta = 1$, so (4.2.3) can be immediately combined with (4.2.2) to get

$$\begin{aligned}
\|f_{X_\varepsilon} - p_X\|_1 &\leq \left\|D^2 p_X\right\|_1 \mathbb{E}_r\left[\int_0^r t\,\eta(t)\,dt\right] \leq \left\|D^2 p_X\right\|_1 \mathbb{E}_r\left[\frac{1}{2}r^2\eta(r)\right] \\
&\leq \eta_{2,n,c_{\mathcal{X}}\varepsilon}\,\frac{n}{2}\left\|D^2 p_X\right\|_1 \varepsilon^2
\end{aligned}$$

When $\delta \in (0, 1]$, note that

$$\begin{aligned}
(1 + c_{\mathcal{X}}|x|)^\delta &\leq (1 + c_{\mathcal{X}}\left|x'\right| + c_{\mathcal{X}}\left|x - x'\right|)^\delta \leq (1 + c_{\mathcal{X}}\left|x'\right| + c_{\mathcal{X}} d_{\mathcal{X}}(x, x'))^\delta \\
&\leq \left[(1 + c_{\mathcal{X}}\left|x'\right|)^\delta + c_{\mathcal{X}}^\delta d_{\mathcal{X}}(x, x')^\delta\right]
\end{aligned}$$

Applying this to (4.2.2) and (4.2.3) we have (using $c_{\mathcal{X}} t \leq c_{\mathcal{X}} R < 1$ and $\|\cdot\|_1 \leq \|\cdot\|_{(\delta)}$)

$$\begin{aligned}
\|f_{X_\varepsilon} - p_X\|_{(\delta)} &\leq \left\|D^2 p_X\right\|_{(\delta)} \mathbb{E}_r\left[\int_0^r t\,\eta(t)\,dt\right] + \left\|D^2 p_X\right\|_1 \mathbb{E}_r\left[\int_0^r (c_{\mathcal{X}} t)^\delta t\,\eta(t)\,dt\right] \\
&\leq 2\left\|D^2 p_X\right\|_{(\delta)} \mathbb{E}_r\left[\int_0^r t\,\eta(t)\,dt\right] \\
&\leq 2\left\|D^2 p_X\right\|_{(\delta)} \mathbb{E}_r\left[\frac{1}{2}r^2\eta(r)\right] \\
&\leq \eta_{2,n,c_{\mathcal{X}}\varepsilon}\left\|D^2 p_X\right\|_{(\delta)} n\varepsilon^2
\end{aligned}$$

□

4.2.2 The $g_{Y_\varepsilon}, g_e,$ and g_o Auxiliary Functions

Define the function $g_{Y_\varepsilon} \in \mathcal{P}(m^N)$, using the shorthand $r = d_{\mathcal{X}}(x, \tilde{x})$, as

$$g_{Y_\varepsilon}(y) := \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi_{n,\varepsilon}^{(R)}(r)\, J_x(\tilde{x})^{-1}\left[1 - \tanh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right)\right] dP_X(\tilde{x})$$

Lemma 4.2.6. *When $R \le \left(\sqrt{2}c_{\mathcal{X}}\right)^{-1}$, we have the following bounds:*

$$\begin{aligned}
\Delta_{\top} &:= \left|\exp\left(-\frac{r^2}{2\varepsilon^2}\delta_{\top}\right) - 1\right| && \le \exp\left(\frac{r^2}{8\varepsilon^2}\right)\frac{c_{\mathcal{X}}^2 r^4}{4\varepsilon^2} \\
\Delta_{\perp} &:= \left|\cosh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_{\perp}\right) - 1\right| && \le \exp\left(\frac{r^2 + |\nu_y|^2}{4\sqrt{2}\varepsilon^2}\right)\frac{c_{\mathcal{X}}^2 r^2 |\nu_y|^2}{8\varepsilon^2} \\
\Delta_{J} &:= |J - 1| && \le \left(1 + \tfrac{1}{4}c_{\mathcal{X}}^2 r^2\right)^{n-1} \\
\Delta_{J^{-1}} &:= \left|J^{-1} - 1\right| && \le \tfrac{1}{4}(n-1)c_{\mathcal{X}}^2 r^2
\end{aligned}$$

Proof. The first two bounds follow from Taylor's theorem, Lemma 4.1.1, and the assumption $c_{\mathcal{X}} r \le 2^{-1/2}$. For the second bound we also use Young's inequality to yield $r|\nu_y| \le \frac{1}{2}\left(r^2 + |\nu_y|^2\right)$. For the last two bounds, start with Corollary 2.4.5. When $x = c_{\mathcal{X}} r \le 2^{-1/2}$ we have, again by Taylor, $\frac{\sinh(x)}{x} \le 1 + \frac{1}{6}\cosh(2^{-1/2})x^2 \le 1 + \frac{1}{4}x^2$. This gives the third bound, and the final bound when combined with $(1+y)^{n-1} \le$ and $(1+y)^{1-n} \ge 1 - (n-1)y + \frac{n(n-1)}{2}y^2$. $\square$

We will use the notation

$$\mathfrak{k}_{k,a,\varepsilon} := \frac{(n+2k-2)!!}{(n-2)!!(1-a)^{k+n/2}}\varepsilon^{2k}$$

Lemma 4.2.7. *If $0 \le a < 1$ and $k \ge 0, k \in \mathbb{Z}$, then $\int_0^\infty e^{\frac{ar^2}{2\varepsilon^2}} r^{2k}\, \chi_{n,\varepsilon}(r)dr = \mathfrak{k}_{k,a,\varepsilon}$.*

Proof. This is easily computed by the change of variable $r \mapsto (1-a)^{1/2} r$. $\square$

Lemma 4.2.8. *For* $R \leq (\sqrt{2}c_{\mathcal{X}})^{-1}$ *and* $\delta \in (0,1]$, *we have*

$$\begin{aligned} \|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_\infty &\leq 2^{n+1}\varepsilon^{-N} \\ \|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_1 &\leq \mathsf{const}(n,N)(c_{\mathcal{X}}\varepsilon)^2 \\ \|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_{(\delta);c_{\mathcal{X}}} &\leq \|P_X\|_{(\delta);c_{\mathcal{X}}} \|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_1 \end{aligned}$$

Remark. f_{Y_ε} and g_{Y_ε} are defined on the tubular neighborhood $\mathcal{U}_R^{\mathcal{X}} \subset \mathbb{R}^N$, and the above norms are with respect to m^N. As an open subset of $\mathbb{R}^N$, $c^o_{\mathcal{U}_R^{\mathcal{X}}} = 0$, but the decay norm is weighted using $a_{\mathcal{U}_R^{\mathcal{X}}} = c_{\mathcal{X}}$ to facilitate conversion to a decay norm on $\mathcal{X}$.

Proof. Using Lemma 4.2.6, $|\tanh t| \leq 1$, and $\left|\frac{t}{\sin t}\right| \leq (1 - \frac{t^2}{6})^{-1} \leq \frac{12}{11}$ when $t = c_{\mathcal{X}} r \leq 2^{-1/2}$, we have the L^∞ bound:

$$\begin{aligned} |f_{Y_\varepsilon} - g_{Y_\varepsilon}| &\leq 2\varphi_{n',\varepsilon}(|\nu_y|) \int \varphi^{(R)}_{n,\varepsilon}(r) \left| e^{-\frac{r^2}{2\varepsilon^2}\delta_\top} \cosh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right) - J_x^{-1} \right| dP_X \\ &\leq 2\varepsilon^{-N} \left\| e^{-\frac{r^2}{2\varepsilon^2}(1 - |\delta_\top| - |\nu_y \cdot \delta_\perp|)} + e^{-\frac{r^2}{2\varepsilon^2}} \frac{c_{\mathcal{X}} r}{|\sin c_{\mathcal{X}} r|} \right\|_\infty \\ &\leq 2\varepsilon^{-N} \left[1 + \left(\tfrac{12}{11}\right)^{n-1} \right] \leq 2^{n+1}\varepsilon^{-N} \end{aligned}$$

For the remaining results we first bound the integrand function more carefully using Lemma 4.2.6:

$$\begin{aligned} \Psi &:= \left| e^{-\frac{r^2}{2\varepsilon^2}\delta_\top} \cosh\left(\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right) - J_x(\tilde{x})^{-1} \right| \\ &\leq \Delta_\top + \Delta_\perp + \Delta_\top \Delta_\perp + \Delta_{J^{-1}} \\ &\leq \frac{c_{\mathcal{X}}^2}{4\varepsilon^2} \left(e^{\frac{r^2}{8\varepsilon^2}} r^4 + \frac{1}{2} e^{\frac{r^2}{5\varepsilon^2}} r^2 \, e^{\frac{|\nu_y|^2}{5\varepsilon^2}} |\nu_y|^2 \right) + \frac{c_{\mathcal{X}}^4}{32\varepsilon^4} e^{\frac{3r^2}{8\varepsilon^2}} r^6 \, e^{\frac{|\nu_y|^2}{5\varepsilon^2}} |\nu_y|^2 + \tfrac{n-1}{4} c_{\mathcal{X}}^2 r^2 \qquad (4.2.4) \end{aligned}$$

Let $\delta \in [0,1]$. We have, by switching order of integration, exploiting the ν_y symmetry of the integrand to eliminate Θ_o, and integrating $dV^n(x)$ with respect to a geodesic normal coordinate system parameterized by $\tau \in T_{\tilde{x}}\mathcal{X}$:

$$\begin{aligned}
\|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_{(\delta);c_\mathcal{X}} &= \int_\mathcal{X}\int_{\mathcal{U}_\mathcal{X}} |f_{Y_\varepsilon}(y) - g_{Y_\varepsilon}(y)|(1 + c_\mathcal{X}|y|)^\delta \, dm^N(y)\, dP_X(\tilde{x}) \\
&\le 2\mathbb{E}_{\tilde{X}} \iint (1 + c_\mathcal{X}|y|)^\delta \varphi^{(R)}_{n',\varepsilon}(|\nu_y|)\varphi^{(R)}_{n,\varepsilon}(r)\, |\Psi(\tilde{x}, x_y, \nu_y)|\, \Theta(x_y, \nu_y)\, dm^{n'}\, dV^n \\
&\le 2\mathbb{E}_{\tilde{X}} \iint (1 + c_\mathcal{X}|y|)^\delta \varphi^{(R)}_{n',\varepsilon}(|\nu_y|)\varphi^{(R)}_{n,\varepsilon}(r)\, |\Psi| \left(1 + \tfrac{1}{4}c_\mathcal{X}^2 r^2\right)^{n-1} \Theta_e\, dm^{n'}\, dm^n
\end{aligned}$$

If $\delta = 0$ we can immediately combine this with (4.2.4), Corollary 2.5.3, and Lemma 4.2.7 to obtain an L^1 bound. While the precise bound is messy, it is easily seen to be $O\big((c_\mathcal{X}\varepsilon)^2\big)$ with bounding constant depending only on n, N.

If $\delta \in (0, 1]$ we have $(1 + c_\mathcal{X}|y|)^\delta \le \big[(1 + c_\mathcal{X}|\tilde{x}|)^\delta + (c_\mathcal{X}|y - \tilde{x}|)^\delta\big]$. Since $|y - \tilde{x}| \le \sqrt{2}R$ we have $\|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_{(\delta);c_\mathcal{X}} \le (1+\|P_X\|_{(\delta);c_\mathcal{X}})\|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_1 \le 2\|P_X\|_{(\delta);c_\mathcal{X}}\|f_{Y_\varepsilon} - g_{Y_\varepsilon}\|_1$.

$\square$

We also have the ν_y-even and odd parts of $g_{Y_\varepsilon} = g_e + g_o$, respectively:

$$\begin{aligned}
g_e &:= \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi^{(R)}_{n,\varepsilon}(r)\, J^{-1}_{x_y}(\tilde{x})\, dP_X(\tilde{x}) = \varphi_{n',\varepsilon}(|\nu_y|) f_{X_\varepsilon}(x_y) \\
g_o &:= \varphi_{n',\varepsilon}(|\nu_y|) \int \varphi^{(R)}_{n,\varepsilon}(r)\, J^{-1}_{x_y}(\tilde{x})\, \tanh\!\left(-\frac{r^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp\right) dP_X(\tilde{x})
\end{aligned}$$

Lemma 4.2.9. *For $R \le \big(\sqrt{2}c_\mathcal{X}\big)^{-1}$ we have*

$$0 \le h_N(g_e) - h_N(g_{Y_\varepsilon}) - \left[\iint (\log g_e)\, g_o \Theta_o\, dm^{n'} dV^n\right] \le \mathsf{const}(n, N)(c_\mathcal{X}\varepsilon)^2$$

Proof. Set $\xi := g_o/g_e$ when $g_e > 0$. Since $|\tanh(s)| < 1$ for all $s \in \mathbb{R}$, $|\xi| < 1$. We will also write $\sigma(y, \tilde{x}) := -\frac{d_\mathcal{X}(x_y,\tilde{x})^2}{2\varepsilon^2}\nu_y \cdot \delta_\perp(\tilde{x})$ for brevity. Since $dm^{n'}(\nu_y)$ is symmetric under $\nu_y \mapsto -\nu_y$, we can symmetrize the entropy integrand using Θ_e and Θ_o:

$$
\begin{aligned}
h_N(g_e) - h_N(g_{Y_\varepsilon}) &= \iint [g_{Y_\varepsilon} \log g_{Y_\varepsilon} - g_e \log g_e] \Theta \, dm^{n'} dV^n \\
&= \frac{1}{2} \iint [(g_e + g_o) \log(g_e + g_o) - g_e \log(g_e)](\Theta_e + \Theta_o) + \\
&\quad + [(g_e - g_o) \log(g_e - g_o) - g_e \log(g_e)](\Theta_e - \Theta_o) \, dm^{n'} \, dV^n \\
&= \iint \psi(\xi) g_e + g_e (\log g_e) \, \xi \Theta_o \, dm^{n'} dV^n \\
&= \iiint \varphi^{(R)}_{n',\varepsilon}(|\nu_y|) \varphi^{(R)}_{n,\varepsilon}(r) \, J^{-1}_{x_y}(\tilde{x}) \psi(\xi) \, dm^{n'} dV^n dP_X + \\
&\quad + \iint (\log g_e) \, g_o \Theta_o \, dm^{n'} dV^n
\end{aligned}
$$

where the auxiliary function ψ is defined on $(-1, 1)$ by

$$
\psi(t) := \frac{\Theta_e + \Theta_o}{2} (1 + t) \log(1 + t) + \frac{\Theta_e - \Theta_o}{2} (1 - t) \log(1 - t)
$$

Since $|\xi| < 1$ and $\Theta_e(x_y, \nu_y) - \Theta_o(x_y, \nu_y) = \Theta(x_y, -\nu_y) > 0$, it is easily checked that $\psi(0) = 0$, $\psi'(t) \geq 0$ for $t > 0$ and ≤ 0 otherwise, and $\psi''(t) \geq 0$. Hence ψ is convex and non-negative. Non-negativity immediately gives

$$
0 \leq h_N(g_e) - h_N(g_{Y_\varepsilon}) - \iint (\log g_e) \, g_o \Theta_o \, dm^{n'} dV^n
$$

By convexity we can apply Jensen's inequality with the probability measure defined (for fixed y) by $d\mu(\tilde{x}) = g_e(y)^{-1} \varphi^{(R)}_{n,\varepsilon}(d_{\mathcal{X}}(x_y, \tilde{x})) \, J^{-1}_{x_y}(\tilde{x}) \, dP_X(\tilde{x})$:

$$
g_e \psi(\xi) = g_e \psi \left[\int \tanh \sigma \, d\mu(\tilde{x}) \right] \leq \int \psi(\tanh \sigma) \varphi^{(R)}_{n,\varepsilon}(r) \, J^{-1}_{x_y}(\tilde{x}) \, dP_X(\tilde{x})
$$

With some algebraic manipulation we have, for all $s \in \mathbb{R}$,

$$\psi(\tanh(s)) = \Theta_e[s \tanh s - \log(\cosh s)] + \Theta_o[s - \tanh(s) \log(\cosh s)]$$

Setting $s = \sigma$, parameterizing x_y in geodesic normal coordinates $\tau \in T_{\tilde{x}}\mathcal{X} \approx \mathbb{R}^n$, and using $|\tanh s| \leq s$ and $0 \leq \log(\cosh s) \leq \frac{1}{2}s^2(\cosh s)$,

$$\begin{aligned} h_N(g_e) - h_N(g_{Y_\varepsilon}) - &\iint (\log g_e)\, g_o \Theta_o \, dm^{n'} dV^n \\ &\leq \iiint \varphi^{(R)}_{n',\varepsilon}(|\nu_y|) \varphi^{(R)}_{n,\varepsilon}(r)\, J^{-1}_{x_y}(\tilde{x})[\sigma \tanh \sigma - \log(\cosh \sigma)]\Theta_e \, dm^{n'} dV^n dP_X \\ &+ \iiint \varphi^{(R)}_{n',\varepsilon}(|\nu_y|) \varphi^{(R)}_{n,\varepsilon}(r)\, J^{-1}_{x_y}(\tilde{x})[\sigma - \log(\cosh \sigma) \tanh \sigma]\Theta_o \, dm^{n'} dV^n dP_X \\ &\leq \iiint \varphi^{(R)}_{n',\varepsilon}(|\nu_y|) \varphi^{(R)}_{n,\varepsilon}(r)\, \eta(r) |\sigma|^2 \, \Theta_e \, dm^{n'} dm^n dP_X \\ &+ \iiint \varphi^{(R)}_{n',\varepsilon}(|\nu_y|) \varphi^{(R)}_{n,\varepsilon}(r)\, \eta(r) \left[|\sigma| \vee \frac{1}{2} |\sigma|^3 (\cosh \sigma) \right] |\Theta_o| \, dm^{n'} dm^n dP_X \end{aligned}$$

Since $|\sigma|^l \leq \left[\frac{r^2}{2\varepsilon^2}\right]^l [c_{\mathcal{X}}|\nu_y|]^l$, and $|\Theta_o| \leq \mathsf{const}(n, N)(c_{\mathcal{X}}|\nu_y|)$, the above terms are all of the form $\mathsf{const}(n, N)(\frac{r^2}{2\varepsilon^2})^l (c_{\mathcal{X}}|\nu_y|)^2$ for $l \in \{1, 2, 3\}$, multiplied by $\phi^{(R)}_{n',\varepsilon}(|\nu_y|)\phi^{(R)}_{n,\varepsilon}(r)$. The result follows by integration. □

Remark. The following scaling properties are straightforward to verify, and will be used in the subsequent two lemmas: if we scale lengths in $\mathbb{R}^N$ by a factor of $a > 0$, the densities $g_{Y_\varepsilon}, g_e, g_o$, etc. scale by a factor of a^{-N}, while densities on $\mathcal{X}$ scale by a^{-n}; The volume elements $dm^{n'}$ and dV^n scale by factors of $a^{n'}$ and a^n, respectively; Jacobian factors Θ_e, Θ_o, J_x, and 1-norms of densities such as $\|g_e\|_1$ are unchanged; $c_{\mathcal{X}}$ scales like a^{-1}; ε scales like a; Entropies change additively as log of volume, e.g. $h_N(q_e) \mapsto h_N(q_e) + \log a^N$. Hence, $(c_{\mathcal{X}}\varepsilon)$ is scale-invariant, as is any difference of two entropies of the same dimension.

Lemma 4.2.10. *Let $R \le \left(\sqrt{2}\, c_{\mathcal{X}}\right)^{-1}$ and $\delta \in (0,1]$. We have*

$$\left| h_N(q_e) - \|g_e\|_1^{-1} h_{V^n}(f_{X_\varepsilon}) - \frac{n'}{2}\log(2\pi e\varepsilon^2) \right| \le \mathsf{const}(n,N)\delta^{-1}\|P_X\|_{(\delta)}(c_{\mathcal{X}}\varepsilon)^2 \log(c_{\mathcal{X}}\varepsilon)^{-1}$$

Proof. Note that $|\|g_e\|_1 - 1| \le \mathsf{const}(n,N)(c_{\mathcal{X}}\varepsilon)^2$. We have

$$\begin{aligned} h_N(g_e) &= \iint \phi^{(R)}_{n',\varepsilon}(|\nu_y|)\Theta_e f_{X_\varepsilon}\left[\log f_{X_\varepsilon}^{-1} + \frac{n'}{2}\log(2\pi\varepsilon^2) + \frac{|\nu_y|^2}{2\varepsilon^2}\right] dm^{n'} dV^n \\ &= h_{V^n}(f_{X_\varepsilon}) + \|g_e\|_1 \frac{n'}{2}\log(2\pi e\varepsilon^2) + \iint \phi^{(R)}_{n',\varepsilon}(|\nu_y|)(\Theta_e - 1) f_{X_\varepsilon}\log f_{X_\varepsilon}^{-1}\, dm^{n'} dV^n \\ &\quad + \iint f_{X_\varepsilon}(x_y)\phi^{(R)}_{n',\varepsilon}(|\nu_y|)[\Theta_e(y) - 1]\, dm^{n'} dV^n \end{aligned}$$

Since $|\Theta_e - 1| \le \mathsf{const}(n,N)(c_{\mathcal{X}}|\nu_y|)^2$, the final term can be bounded in absolute value by $\mathsf{const}(n,N)(c_{\mathcal{X}}\varepsilon)^2$. The result follows if we can bound the term $\iint \phi^{(R)}_{n',\varepsilon}(|\nu_y|)(\Theta_e - 1) f_{X_\varepsilon}\log f_{X_\varepsilon}^{-1}\, dm^{n'} dV^n$, which is bounded as follows:

$$\begin{aligned} \iint \phi^{(R)}_{n',\varepsilon}(|\nu_y|)|\Theta_e - 1| f_{X_\varepsilon}\left|\log f_{X_\varepsilon}^{-1}\right| dm^{n'} dV^n &\le \mathsf{const}(n,N)(c_{\mathcal{X}}\varepsilon)^2 \int f_{X_\varepsilon}\left|\log f_{X_\varepsilon}^{-1}\right| dV^n \\ &\le \mathsf{const}(n,N)(c_{\mathcal{X}}\varepsilon)^2\left[h_{V^n}(f_{X_\varepsilon}) + 2\int_{f_{X_\varepsilon}>1} f_{X_\varepsilon}\log f_{X_\varepsilon}\, dV^n\right] \\ &\le \mathsf{const}(n,N)(c_{\mathcal{X}}\varepsilon)^2\left[h_{V^n}(f_{X_\varepsilon}) + 2\|f_{X_\varepsilon}\|_1 \log^+\|f_{X_\varepsilon}\|_\infty\right] \end{aligned}$$

By definition, $\|f_{X_\varepsilon}\|_\infty \le \left[\frac{2^{-1/2}}{\sin\left(2^{-1/2}\right)}\right]^{(n-1)}\left(2\pi\varepsilon^2\right)^{-n/2} \le \varepsilon^{-n}$, so this bound is finite. Furthermore, since the quantities $h_N(g_e) - \|g_e\|_1^{-1} h_{V^n}(f_{X_\varepsilon}) - \frac{n'}{2}\log(2\pi e\varepsilon^2)$, $\|f_{X_\varepsilon}\|_1$, and $(c_{\mathcal{X}}\varepsilon)$ are scale-invariant, we are free to set any scale convenient for bounding. Scale lengths by $a = \varepsilon^{-1}$, so $\tilde{\varepsilon} = 1$ and $\tilde{c}_{\mathcal{X}} = c_{\mathcal{X}}\varepsilon$. This removes the $\log^+\|f_{X_\varepsilon}\|_\infty$ piece of the bound, and applying Corollary 3.4.3 gives the bound

$$h_{V^n}(f_{X_{\bar{\varepsilon}}}) \le e^3\Big[\big(2 + N\delta^{-1}\big) \vee \log(\kappa_{N,n}\omega_n) + n\log(c_{\mathcal{X}}\varepsilon)^{-1}\Big]\|f_{X_\varepsilon}\|_{(\delta)}$$

where we used the scale-invariance of $\|f_{X_\varepsilon}\|_{(\delta)}$ and the fact that $\log\|f_{X_\varepsilon}\|_{(\delta)}^{-1} \le |1 - \|f_{X_\varepsilon}\|_1| \le \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2$. The result follows by noting that $h_N(q_e) = \|g_e\|_1^{-1} h_N(g_e) + \log\|g_e\|_1$ and $\log\|g_e\|_1 \le \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2$. $\square$

Lemma 4.2.11. *Let $R \le \big(\sqrt{2}\,c_{\mathcal{X}}\big)^{-1}$. Set $c_e := \|g_e\|_1$, $c_o := \|g_{Y_\varepsilon}\|_1 - c_e \equiv \int g_o\, dm^N$. Define the probability densities $q_{Y_\varepsilon} := (c_e + c_o)^{-1} g_{Y_\varepsilon}$ and $q_e := c_e^{-1} g_e$. We have*

$$|h_N(q_e) - h_N(q_{Y_\varepsilon})| \le \mathsf{const}(n, N)\delta^{-1}\|P_X\|_{(\delta)}\big[(c_{\mathcal{X}}R)^2 + (c_{\mathcal{X}}\varepsilon)^2\big]\log(c_{\mathcal{X}}\varepsilon)^{-1}$$

Proof. Set $c_2 := \iint |g_o\Theta_o|\, dm^{n'}\, dV^n$. We have, using anti-symmetry, $|\tanh(s)| \le s$, $|\sigma| \le \frac{r^2}{2\varepsilon^2} c_{\mathcal{X}}|\nu_y|$, and $|\Theta_o| \le \mathsf{const}(n, N)c_{\mathcal{X}}|\nu_y|$:

$$\begin{aligned}
|c_o| &= \left|\iint g_o\Theta\, dm^{n'}\, dV^n\right| = \left|\iint g_o\Theta_o\, dm^{n'}\, dV^n\right| \le c_2 \\
c_2 &\le \iiint \varphi_{n',\varepsilon}^{(R)}(|\nu_y|)\varphi_{n,\varepsilon}^{(R)}(r)\frac{r^2}{2\varepsilon^2}c_{\mathcal{X}}|\nu_y|J_x^{-1}(\tilde{x})|\Theta_o(x_y, \nu_y)|\, dm^{n'}\, dV^n\, dP_X \\
&\le \mathsf{const}(n, N)\iiint \left[\chi_{n,\varepsilon}^{(R)}\frac{r^2}{\varepsilon^2}\eta(r)\right]\left[\varphi_{n',\varepsilon}(|\nu_y|)(c_{\mathcal{X}}|\nu_y|)^2\right] dr\, dm^{n'}\, dP_X \\
&\le \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2
\end{aligned}$$

We also have, from previous estimates, $|c_e - 1| \le \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2$.

By the definition of entropy, when $\|q\|_1 = 1$, $h(c\,q) = c\,h(q) - c\log c$, so

$$h_N(g_e) - h_N(g_{Y_\varepsilon}) = c_e\big[h_N(q_e) - h_N(q_{Y_\varepsilon}) + \log\big(1 + c_e^{-1}c_o\big)\big] + c_o[\log c_e - h_N(q_e)]$$

Thus Lemma 4.2.9 can be written

$$
\begin{aligned}
0 \leq & h_N(q_e) - h_N(q_{Y_\varepsilon}) + \log\left(1 + \frac{c_o}{c_e}\right) + \frac{c_o}{c_e + c_o} \log c_e - \frac{c_o}{c_e + c_o} h_N(q_e) \\
& + \frac{1}{c_e + c_o} \iint \left[\log g_e^{-1}\right] g_o \Theta_o \, dm^{n'} \, dV^n \leq \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2
\end{aligned}
$$

Note that, since $\log g_e^{-1} = \frac{n'}{2} \log(2\pi\varepsilon^2) + \frac{|\nu_y|^2}{2\varepsilon^2} + \log f_{X_\varepsilon}^{-1}$,

$$
\begin{aligned}
\iint \left[\log g_e^{-1}\right] g_o \Theta_o \, dm^{n'} \, dV^n = c_o \frac{n'}{2} \log(2\pi\varepsilon^2) + \iint \frac{|\nu_y|^2}{2\varepsilon^2} g_o \Theta_o \, dm^{n'} \, dV^n + \\
+ \iint \left[\log f_{X_\varepsilon}^{-1}\right] g_o \Theta_o \, dm^{n'} \, dV^n
\end{aligned}
$$

Furthermore, $\left| \iint \frac{|\nu_y|^2}{2\varepsilon^2} g_o \Theta_o \, dm^{n'} \, dV^n \right| \leq \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2$, which can be shown by the same method used above to bound c_2. Combining these inequalities, the bound of c_o, and the $h_N(q_e)$ estimate (Lemma 4.2.10) we have, after some algebra, and setting $g_{X_\varepsilon} := c_e^{-1} f_{X_\varepsilon}$, so that $\left|h_{V^n}(g_{X_\varepsilon}) - c_e^{-1} h_{V^n}(f_{X_\varepsilon})\right| = \frac{\|f_{X_\varepsilon}\|_1}{c_e} \log c_e \leq \mathsf{const}(n, N)(c_{\mathcal{X}}\varepsilon)^2$,

$$
\begin{aligned}
|h_N(q_e) - h_N(q_{Y_\varepsilon})| \leq \left| \iint \left[\log g_{X_\varepsilon}^{-1} - h_{V^n}(g_{X_\varepsilon})\right] g_o \Theta_o \, dm^{n'} dV^n \right| \\
+ \mathsf{const}(n, N) \delta^{-1} \|P_X\|_{(\delta)} (c_{\mathcal{X}}\varepsilon)^2 \log(c_{\mathcal{X}}\varepsilon)^{-1}
\end{aligned}
$$

Since $|g_o \Theta_o| \leq \mathsf{const}(n, N) R^2 \left(c_{\mathcal{X}} |\nu_y| \varepsilon^{-1}\right)^2 \varphi_{n',\varepsilon}(|\nu_y|) g_{X_\varepsilon}(x_y)$, we have

$$
\begin{aligned}
\left| \iint \left[\log g_{X_\varepsilon}^{-1} - h_{V^n}(g_{X_\varepsilon})\right] g_o \Theta_o \, dm^{n'} dV^n \right| \leq \\
\leq \mathsf{const}(n, N)(c_{\mathcal{X}} R)^2 \int \left|\log g_{X_\varepsilon}^{-1} - h_{V^n}(g_{X_\varepsilon})\right| g_{X_\varepsilon} \, dV^n
\end{aligned}
$$

Finally, note that

$$\begin{aligned}\int \left|\log g_{X_\varepsilon}^{-1} - h_{V^n}(g_{X_\varepsilon})\right| g_{X_\varepsilon}\, dV^n &= \int \left[\log g_{X_\varepsilon}^{-1} - h_{V^n}(g_{X_\varepsilon})\right] g_{X_\varepsilon}\, dV^n + \\ &\quad + \int\limits_{\{g_{X_\varepsilon} > \exp(-h_{V^n}(g_{X_\varepsilon}))\}} 2[\log g_{X_\varepsilon} + h_{V^n}(g_{X_\varepsilon})] g_{X_\varepsilon}\, dV^n \\ &= \int\limits_{\{g_{X_\varepsilon} > \exp(-h_{V^n}(g_{X_\varepsilon}))\}} 2[\log g_{X_\varepsilon} + h_{V^n}(g_{X_\varepsilon})] g_{X_\varepsilon}\, dV^n \\ &\leq 2\|g_{X_\varepsilon}\|_1 [h_{V^n}(g_{X_\varepsilon}) - \log\|g_{X_\varepsilon}\|_\infty]\end{aligned}$$

We have $\|g_{X_\varepsilon}\|_\infty \leq c_e^{-1}\varepsilon^{-n}$. As in the previous lemma, our expressions are scale-invariant, and we obtain a bound

$$\int \left|\log g_{X_\varepsilon}^{-1} - h_{V^n}(g_{X_\varepsilon})\right| g_{X_\varepsilon}\, dV^n \leq \mathsf{const}(n, N) \|P_X\|_{(\delta)} \log(c_{\mathcal{X}}\varepsilon)^{-1}$$

which completes the proof. □

4.3 Power-Constrained Channel Capacity

For $2 \leq n \leq N$, let Ω be a compact $(n-1)$-dimensional differentiable submanifold of S^{N-1}, i.e., a compact submanifold of $\mathbb{R}^N$ such that $|\omega| = 1$ for all $\omega \in \Omega$. Define $\mathcal{X} = \Omega \times \mathbb{R}^+ = \{r\omega : r > 0, \omega \in \Omega\}$ and a channel $\mathcal{X} \to \mathcal{Y} = \mathbb{R}^N$ by AWGN of average power ε^2. We impose the average power constraint $\mathbb{E}|X|^2 \leq n\mathsf{P}$. Define $\mathsf{SNR} := \frac{\mathsf{P}}{\varepsilon^2}$.

Theorem 4.3.1 (Average Power-Constrained Asymptotic Channel Capacity)**.** *As* $\mathsf{SNR} \to \infty$*, the capacity of the channel described above is asymptotically given by*

$$\mathsf{Cap}(\mathsf{SNR}) \approx \frac{n}{2}\log(1 + \mathsf{SNR}) + \log\frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})}$$

For $\mathsf{SNR} > 1$ *the rate of convergence is bounded from above and below as follows:*

$$-\mathsf{const}(N, c_\Omega)\left(\frac{1}{\mathsf{SNR}}\right)^{\frac{n}{n+2}} \log^3(\mathsf{SNR}) \leq \mathsf{Cap}(\mathsf{SNR}) - \left[\frac{n}{2}\log(1+\mathsf{SNR}) + \log\frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})}\right]$$
$$\leq \mathsf{const}(N, c_\Omega)\left(\frac{1}{\mathsf{SNR}}\right)^{\frac{N}{N+2}} \log^3(\mathsf{SNR})$$

To prove the theorem we first maximize $h_{V^n}(X)$:

Lemma 4.3.2. *$h_{V^n}(X)$ is maximized when $\mathsf{P}^{-1/2}|X|$ is distributed as χ_n and $\hat{X} := X/|X|$ is distributed uniformly over Ω, independent of $|X|$. The achieved maximum entropy is*

$$h_{V^n}(X) = \frac{n}{2}\log(2\pi e\mathsf{P}) + \log\frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})}$$

Proof. Expressing entropy in the $\left(|X|, \hat{X}\right)$ polar coordinates:

$$\begin{aligned} h(X) &= h_{dr\times V^{n-1}}\left(P_{|X|,\hat{X}}\right) + \mathbb{E}\log|X|^{n-1} \\ &= \left[h(|X|) + \mathbb{E}\log|X|^{n-1}\right] + \left[h_{V^{n-1}}\left(\hat{X}\right)\right] - I\left(|X|;\hat{X}\right) \end{aligned}$$

The sum is maximized when $P_{|X|} \perp P_{\hat{X}}$ and the bracketed terms are individually maximized. By Lemma 3.1.1, $h_{V^{n-1}}(\hat{X})$ has maximum $\log V^{n-1}(\Omega)$, achieved by the uniform pdf $\left[V^{n-1}(\Omega)\right]^{-1} dV^{n-1}$ on Ω. To maximize the first term, note that in the special case of $N = n$ and $\Omega = S^{n-1}$, $\mathcal{X} = \mathbb{R}^n$, and we maximize $h_{V^n}(X)$ with $X_g \sim \mathcal{N}(0, \mathsf{P}I_n)$. In this case in polar coordinates we have $P_{|X_g|} \perp P_{\hat{X}_g}$, $P_{\hat{X}_g}$ uniform on S^{n-1}, and $P_{|X_g|/\mathsf{P}^{1/2}} \sim \chi_n$. This gives $h(X_g) = \frac{n}{2}\log(2\pi e\mathsf{P}) = h(\chi_n) + \mathbb{E}_{\chi_n}\log|X_g|^{n-1} + \log\left|S^{n-1}\right|$, so for any $P_{|X|}$ satisfying $\mathbb{E}|X|^2 \leq n\mathsf{P}$ we have $h(|X|) + \mathbb{E}\log|X|^{n-1} \leq \frac{n}{2}\log(2\pi e\mathsf{P}) - \log V^{n-1}\left(S^{n-1}\right)$, with this maximum achieved when $\mathsf{P}^{-1/2}|X|$ obeys the χ_n distribution. $\square$

Proof of Theorem 4.3.1. We assume $\mathsf{SNR} > 1$. By overall scale-invariance of SNR and mutual information, we may assume without loss of generality that $\mathsf{P} = 1$. Since $\Omega \subset S^{N-1}$ and is compact, it is uniform with $1 < c_\Omega < \infty$. Let $\alpha \in (0,1)$ be number we will choose later, and define

$$\mathcal{X}^{(0)} := \{X \in \mathcal{X} \colon |X|^2 \geq n(\mathsf{SNR})^{-\alpha}\}$$
$$\mathcal{X}^{(1)} := \{X \in \mathcal{X} \colon |X|^2 < n(\mathsf{SNR})^{-\alpha}\} \equiv \mathcal{X} \setminus \mathcal{X}^{(0)}$$
$$\mathcal{Y}^{(0)} := \{Y \in \mathbb{R}^N \colon |Y|^2 \geq n\big(\mathsf{SNR}^{-\alpha}\big)\}$$
$$\mathcal{Y}^{(1)} := \mathbb{R}^N \setminus \mathcal{Y}^{(0)}$$

Also define the random variable $K \in \{0,1\}$ so that $K = k \iff X \in \mathcal{X}^{(k)}$, and put $a_k := P_K(k)$. Set $P_{X^{(k)}} := P_{X|K=k}$, and similarly for $Y_\varepsilon^{(k)}, X_\varepsilon^{(k)}$.

To prove the capacity estimate we need only develop the corresponding upper and lower bounds on $h_N(Y_\varepsilon)$. These estimates each have pieces corresponding to $\mathcal{X}^{(0)}$, the "nice" piece, and $\mathcal{X}^{(1)}$, where the uniformity assumptions break down.

First we look at the nice piece. Since the uniformity bounds scale like inverse length, the (non-compact) submanifold-with-boundary $\mathcal{X}^{(0)}$ is uniform with $c_0 \equiv c_{\mathcal{X}^{(0)}} = n^{-1/2}(\mathsf{SNR})^{\alpha/2}c_\Omega \iff (c_0\varepsilon)^2 = \big(n^{-1}c_\Omega^2\big)(\mathsf{SNR})^{\alpha-1}$. Also note that, if $\delta \in (0,1]$ and $P_{X^{(0)}} \in \hat{\mathcal{P}}(\mathcal{X}^{(0)})$ satisfies $\mathbb{E}|X|^2 \leq n$ then, by Jensen's inequality, $\|P_{X^{(0)}}\|_{(\delta);c_0} \leq 1 + (\sqrt{n}\, c_0)^\delta \leq 1 + c_\Omega^\delta\, (\mathsf{SNR})^{\alpha\delta/2} \leq 2c_\Omega^\delta(\mathsf{SNR})^{\alpha\delta/2}$. Theorem 4.2.1 applies for sufficiently high SNR. Applying the previous observations and using $\gamma > 0, t \geq 1 \implies \log t = \gamma^{-1}\log t^\gamma \leq \gamma^{-1}t^\gamma$ to convert logs to exponents, we have (for $\gamma \in (0,1)$ to be specified later),

$$\left|h(Y_\varepsilon^{(0)}) - \frac{n'}{2}\log\bigl(2\pi e\varepsilon^2\bigr) - h_{V^n}(f_{X_\varepsilon^{(0)}})\right| \le \mathsf{const}(N)\delta^{-1}\|P_X\|_{(\delta)}\Bigl[(c_0\varepsilon)\log(c_0\varepsilon)^{-1}\Bigr]^2$$
$$\le \mathsf{const}(N)\delta^{-1}\|P_X\|_{(\delta)}\Bigl[\gamma^{-1}(c_0\varepsilon)^{1-\gamma}\Bigr]^2$$
$$\le \mathsf{const}(N)\frac{c_\Omega^{2+\delta-2\gamma}}{\delta\gamma^2}\left(\frac{1}{\mathsf{SNR}}\right)^{(1-\gamma)(1-\alpha)-\delta\alpha/2}$$

To obtain an upper bound on $h_N(Y_\varepsilon)$, hence also capacity, note that

$$\begin{aligned} h_N(Y_\varepsilon) &\le h_N(Y_\varepsilon, K) = h_N(Y_\varepsilon|K) + H(K) \\ &\le a_0\, h_N(Y_\varepsilon^{(0)}) + \Bigl[a_1\, h_N(Y_\varepsilon^{(1)}) + H(K)\Bigr] \end{aligned}$$

It is a straightforward exercise in calculus to show that that bracketed term is maximized when $a_1 = \Bigl[1+\exp\Bigl(h(Y_\varepsilon^{(1)})\Bigr)\Bigr]^{-1}$, with maximal value $\log\Bigl[1+\exp\Bigl(-h_N(Y_\varepsilon^{(1)})\Bigr)\Bigr]$. Since $\bigl|X^{(1)}\bigr|^2 \le n(\mathsf{SNR})^{-\alpha}$, $h_N(Y_\varepsilon^{(1)}) \le \frac{N}{2}\log\bigl(2\pi e n\mathsf{SNR}^{-\alpha}\bigr)$, and so

$$h_N(Y_\varepsilon) \le a_0\, h_N(Y_\varepsilon^{(0)}) + \mathsf{const}(N)\mathsf{SNR}^{-N\alpha/2}$$

Applying our estimate of $h_N(Y_\varepsilon^{(0)})$ and Lemma 4.3.2, we have (once SNR is large enough to guarantee $h_N(Y_\varepsilon^{(0)}) \ge 0$ so we may replace a_0 with 1 in our bound)

$$\begin{aligned} I(X;Y_\varepsilon) \le \frac{n}{2}\log(\mathsf{SNR}) + \log\frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})} &+ \mathsf{const}(N,c_\Omega)\delta^{-1}\gamma^{-2}\left(\frac{1}{\mathsf{SNR}}\right)^{(1-\gamma)(1-\alpha)-\delta\alpha/2} \\ &+ \mathsf{const}(N)\left(\frac{1}{\mathsf{SNR}}\right)^{N\alpha/2} \end{aligned}$$

Setting $\alpha = \frac{2}{N+2}, \gamma = \delta = [1+\log\mathsf{SNR}]^{-1}$ completes the upper bound on capacity.

For the lower bound, we will use the following twice: if $\int g\, d\mu \le m$, $0 \le g \le M$, and $\mu(\text{supp}(g)) \leq V$, then $-m \log M \leq h_\mu(g) \leq m \log \frac{V}{m}$. Set P_X as in Lemma 4.3.2. The second part of Theorem 4.2.1, for well-behaved P_X, now also applies, so $h_n(f_{X_\varepsilon})$ may be replaced by $h_n(X^{(0)})$ for the $\mathcal{Y}^{(0)}$ piece of the $h_N(Y_\varepsilon)$ estimate. Since the χ pdf is bounded we also have $a_1 \equiv P_X(\mathcal{X}^{(1)}) \leq \mathsf{const}(N)\mathsf{SNR}^{-n\alpha/2}$, so the exclusion of the $\mathcal{X}^{(1)}$ piece in our $h_n(P_X)$ estimate (omitted since Theorem 4.2.1 does not apply) incurs an error bound $|h_n(1_{\mathcal{X}^{(1)}} P_X)| \leq \mathsf{const}(N)\mathsf{SNR}^{-n\alpha/2} \log(\mathsf{SNR})$. Additionally, since $p_{Y_\varepsilon} \leq \mathsf{const}(N)\mathsf{SNR}^{N/2}$, the $\mathcal{Y}^{(1)}$ component of $h_N(Y_\varepsilon)$ may be bounded similarly using our observation, as $\mathsf{const}(N)\mathsf{SNR}^{-N\alpha/2}$. Combined we have the lower bound estimate

$$I(X;Y_\varepsilon) \geq \frac{n}{2}\log(\mathsf{SNR}) + \log \frac{V^{n-1}(\Omega)}{V^{n-1}(S^{n-1})} + \mathsf{const}(N, c_\Omega)\delta^{-1}\gamma^{-2}\left(\frac{1}{\mathsf{SNR}}\right)^{(1-\gamma)(1-\alpha)-\delta\alpha/2}$$
$$+\mathsf{const}(N)\mathsf{SNR}^{-n\alpha/2}\log(\mathsf{SNR}) + \mathsf{const}(N)\left(\frac{1}{\mathsf{SNR}}\right)^{n\alpha/2}$$

Setting $\alpha = \frac{2}{n+2}, \gamma = \delta = [1 + \log \mathsf{SNR}]^{-1}$ completes the lower bound, and the proof.

□

5

Application to Radar Communication Channel

In this chapter we examine in-depth the application of our main theorems to radar and communications system spectrum sharing. This topic has been the subject of much recent research from a variety of perspectives. For example, in [1] a joint radar and communications channel is abstracted as a unified hybrid channel with rates of both standard information transmission and of “radar estimation information” for an existing, known radar target. Theoretical bounds on joint rates are developed in this framework. While this approach is interesting, the actual radar operation has been abstracted to the point that no path towards implementation can be suggested by the research. The the radar hardware, transmit waveforms, and signal processing algorithms required to even approach the theoretical bound are all abstracted away, remaining completely unaddressed.

Other recent work has focused on practical implementations. Perhaps the simplest and most straightforward approach to spectrum-sharing is the use of time/frequency hopping to prevent cross-interference, as explored in [8]. However, this technique precludes any mutually-beneficial cooperation between the radar and communication systems, such as allowing radar transmissions to act as an amplifier and repeater for communication relays. For cooperation the radar waveform must be allowed to encode information by varying its transmit waveform. One approach is to designate an existing family of radar waveforms as the coding alphabet, as is proposed in [6] for so-called *Oppermann sequences*. However, fixing an ad-hoc family of waveforms is sub-optimal, particularly in the high-SNR regime of interest in this dissertation.

By applying the results of Chapter 4, our approach lies somewhere between these two extremes. In principle any family of radar waveforms, chosen for the desired application, may be analyzed, and the corresponding high-SNR channel capacity, as a function of the chosen waveform performance metrics, may be computed numerically. In this chapter we present an extended study of a relatively simple, but realistic radar system model.

5.1 Radar and Signal Processing Background

Consider a stationary narrowband radar transmitting the waveform $s(t)$ supported on $[0, T]$. We assume $\hat{S}(f) \equiv \int_{-\infty}^{\infty} s(t)e^{-2\pi itf}\,dt$ is concentrated in $[f_0 - \frac{W}{2}, f_0 + \frac{W}{2}]$

where f_0 is the carrier frequency and W the bandwidth. In this section, we normalize to $\|s\|_2 = 1$ and AWGN of power spectral density ε^2.

Compared to $s(t)$, return scatter from a target at range d and radial velocity $\dot{d}$ relative the radar exhibits a time-delay $\tau = \frac{2d}{c}$ and (narrowband) Doppler shift $\sigma = -\frac{2\dot{d}}{c} f_0$, where c is the speed of light. In addition, there is an overall scale factor due to losses, and an overall phase shift which is typically modeled as a uniform random variable on $[0, 2\pi]$, chosen independently for each target. Note that the target return Doppler shift is negligible in the regime $|\sigma T| \ll 1$, or equivalently, when $\dot{d}T \ll \lambda_0$, where λ_0 is the wavelength corresponding to the center frequency. For many common radars (e.g. air traffic control radars), this assumption holds true for all realistic $\dot{d}$. In this paper we will neglect Doppler shift for simplicity, although it can be accounted for, if necessary, using similar techniques.

If we define the time-shifted variant of s

$$s_\tau(t) \equiv s(t - \tau)$$

then return from a scatterer can be written $c \cdot s_\tau(t)$ for some $c \in \mathbb{C}$ with uniform random phase. The radar receives signal $r(t) = \sum_k c_k s_{\tau_k}(t) + N(t)$, a superposition of possible scatterers summed over a discretized, finite set of possible scatterer "bins", plus the random AWGN term $N(t)$. It is typical to process this through matched filters for the s_τ's of interest:

$$\begin{aligned}
f(\tau) &= \int r(t)\overline{s_\tau}(t)\,dt \\
&= \sum_k c_k \int s(t-\tau_k)\overline{s}(t-\tau)\,dt + \int N(t)\overline{s_\tau}(t)\,dt \\
&= \sum_k c_k \int s(t)\overline{s}(t+\Delta_k)\,dt + N_1
\end{aligned}$$

where $\Delta_k \equiv \tau_k - \tau$ and N_1 is complex-normal of variance ε^2. If $|f(\tau)| \gg \varepsilon$ we conclude that a target is present at time-delay τ. It is well-known that the matched filter optimizes SNR among linear filters in the case of a single target with $\Delta = 0$ and AWGN. However, $f(\tau)$ may still be large in the absence of a target at τ, if there is a large target at $\tau_k \neq \tau$ and $\int s(t)\overline{s}(t+\Delta_k)\,dt$ is not very small. Therefore, it is desirable to ensure low "sidelobes" in Δ.

If we restrict our signal processing to a matched filter, the low sidelobe requirement must be enforced by requiring strict cross-correlation properties on s, which limits information capacity considerably. Instead we allow more flexibility in transmit signal and attempt to adapt the filter to the chosen signal. To maintain processing time similar to a matched filter, we consider an arbitrary normalized time-independent linear filter defined for each τ by $w \in L^2$ with $\|w\|_2 = 1$, as

$$\begin{aligned}
f(\tau) &= \langle w_\tau, r\rangle \\
&= \sum_k c_k \langle w_\tau, s_{\tau_k}\rangle + \langle w_\tau, N\rangle \\
&= \sum_k c_k \psi_w(\Delta_k) + N_1
\end{aligned}$$

where $\psi_w(\tau) \equiv \langle w, s_\tau\rangle$.

The assumption of independent uniformly random phases of the c_k and N implies the following simple expression for $\mathbb{E}|f(\tau)|^2$ in which all cross terms vanish:

$$\mathbb{E}|f(\tau)|^2 = \lambda_0 \sum_{k=1}^{K} |\psi_w(\Delta_k)|^2 \frac{1}{K} + \varepsilon^2$$

where we have assumed k ranges over all possible "bins", targets are equally likely to appear in any of the K bins, and a value $\lambda_0 \equiv \mathbb{E}|c_k|^2$ is specified based on judgment of the frequency and scattering characteristics of typical targets tracked by the radar in question. To quantify filter effectiveness it is reasonable to define *signal* using the piece of $\mathbb{E}|f(\tau)|^2$ contributed by the desired target, conditioned on that target appearing at $\tau = 0$. Up to a multiple of λ_0, this is simply

$$\mathsf{S}_s(w) = |\langle s, w \rangle|^2$$

To define *interference power*, first note that for s (approximately) band-limited to $[f_0 - \frac{W}{2}, f_0 + \frac{W}{2}]$, $\psi(\tau) = \langle s_\tau, w \rangle$ is also band-limited to the same frequency interval, so by the uncertainty principle its localization in τ is (approximately) bounded below by the characteristic width W^{-1}. Therefore, any w achieving reasonable gain on a target at $\tau = 0$ will also achieve some gain on targets $|\tau| \lesssim T^{-1}$, and in defining interference, it is desirable to exclude contributions from targets within this "guard region" (In practice, the guard region width may be tweaked to adjust the trade-off between resolution and interference suppression; our choice is representative). Average interference is thus defined (again up to a multiple of λ_0) by conditioning on targets outside the guard region, as

$$\mathsf{I}_s(w) = \sum_{|\Delta_k| \leq T^{-1}} |\psi_w(\Delta_k)|^2 \frac{1}{K'}$$

where K' is the number of bins outside the guard region. However, we will be more interested in the interference piece alone in what follows: For any $\lambda \in [0,1]$ we can define

$$R_\lambda(t_1, t_2) := (1-\lambda) \sum_{|\Delta_k| \leq T^{-1}} \overline{s}(t_1 - \Delta_k) s(t_2 - \Delta_k) \frac{1}{K'} + \lambda \delta(t_1 - t_2)$$

and its associated quadratic form $\mathcal{R}_\lambda$:

$$\langle w, \mathcal{R}_\lambda w \rangle = \iint R_\lambda(t_1, t_2) w(t_1) \overline{w}(t_2)\, dt_1\, dt_2$$

Note that we can now write $\mathsf{I}_s(w) = \langle w, \mathcal{R}_0 w \rangle$. For $\lambda > 0$ the Hermitian linear map $\mathcal{R}_\lambda \colon L^2 \to L^2$ is bounded and positive-definite, with eigenvalues bounded below by λ. Hence it is invertible, with a well-defined Hermitian, positive-definite, invertible square-root.

We write *signal-to-interference-plus-noise ratio* (SINR) associated to w and noise-loading parameter $\lambda \in [0,1]$ is

$$\mathsf{SINR}_s(\lambda, w) := \frac{|\langle s, w \rangle|^2}{\langle w, \mathcal{R}_\lambda w \rangle} \tag{5.1.1}$$

In particular, we define the *signal-to-interference ratio* associated to w by $\mathsf{SIR}_s(w) = \mathsf{SINR}_s(0, w)$. It is well-known, (e.g. by changing basis via $\tilde{w} = \mathcal{R}^{-\frac{1}{2}} w$, or using calculus of variations) that ratios of the form 5.1.1 are maximized among unit-normalized filters w, by

$$w_{\text{opt}} = \frac{\mathcal{R}_\lambda^{-1} s}{\left\| \mathcal{R}_\lambda^{-1} s \right\|_2} \tag{5.1.2}$$

The corresponding maximal $\mathsf{SINR}_s(\lambda, w)$ thus achieved is easily computed to be $\left\langle s, \mathcal{R}_\lambda^{-1} s \right\rangle$.

Depending on the specific operational needs of the radar, an acceptable transmit waveform will often be desired to not only satisfy constraints on the achievable optimal-filter SIR, but "double constraints" on the simultaneously achievable optimal-filter values of SIR and S. That is, we may require that every transmitted waveform possesses an associated range filter which simultaneously achieves both a minimal SIR and a minimal S. The following notations express this double waveform filter constraint from two perspectives. The first seeks to maximize SIR while requiring a minimum S, while the second reverses the relationship:

$$\mathsf{SIR}_\alpha(s) := \sup_{\{\|w\|_2 = 1, \mathsf{S}_s(w) \geq \alpha\}} \mathsf{SIR}_s(w)$$

$$\mathsf{S}_\beta(s) := \sup_{\{\|w\|_2 = 1, \mathsf{SIR}_s(w) \geq \beta\}} \mathsf{S}_s(w)$$

$\mathsf{SIR}_\alpha(s)$ is defined for all $\alpha \leq 1$. If $\mathcal{R}_0$ is invertible then $\mathsf{SINR}_s(0, w) < \infty$, and this is the largest value of β for which $\mathsf{S}_\beta(s)$ is defined.

Thus, if we demand only radar waveforms that may be signal processed to achieve $\mathsf{S}_s \geq \alpha$ and $\mathsf{SIR}_s \geq \beta$ with the same filter, this is equivalent to requiring $\Phi_1(s) \equiv \mathsf{SIR}_\alpha(s) \geq \beta \equiv c_1$. Alternately, this is equivalent to instead requiring $\Phi_1(s) \equiv \mathsf{S}_\beta(s) \geq \alpha \equiv c_1$. Below we will use both variations as indicated.

Finally, we define a commonly-used radar waveform which we will use to benchmark the performance of our alphabet waveforms. For a specified transmit time T and bandwidth W, a (symmetric, baseband) *chirp* is a complex waveform of constant amplitude and quadratic phase progression on $[0, T]$:

$$s_{\text{chirp}}(t) = c\, e^{i\pi WT(t/T-1/2)^2}$$

The chirp is characterized by instantaneous frequency varying linearly in time and sweeping out the specified bandwidth. It is commonly used as a radar transmit waveform because it is both simple to generate in hardware and achieves very good SIR under a matched filter (i.e. taking $w{=}s_{\text{chirp}}$). We denote this SIR by β_{chirp}.

5.2 Finite Dimensional Manifold Approximation

To represent our signal space in $\mathbb{R}^N$ for a finite N, we note that s is time-limited and approximately band-limited. Capacity is unaffected by shifting all s to baseband, so WLOG we take s band-limited to $[-\frac{W}{2}, \frac{W}{2}]$. A natural basis to consider are the so-called *prolate spheroidal wave functions*, $\{\phi_k\}_{k=0}^{\infty}$: Let $\iota\colon L^2([0,T]) \to L^2(\mathbb{R})$ and $\pi\colon L^2(\mathbb{R}) \to L^2([-\frac{W}{2}, \frac{W}{2}])$ be the inclusion map and projection (i.e. restriction) map, respectively. If we define $P = \pi \circ \mathcal{F} \circ \iota\colon L^2([0,T]) \to L^2([-\frac{W}{2}, \frac{W}{2}])$, then the $\{\phi_k\}$ are the orthonormal eigenbasis associated with the positive-definite, self-adjoint, compact operator $P^*P\colon L^2([0,T]) \to L^2([0,T])$, guaranteed to exist by the spectral theorem. They can be taken to be real-valued since it is easy to check that P^*P is invariant under conjugation. By convention, we order the ϕ_k's by decreasing eigenvalues. By definition, the eigenvalue λ_k always lies in $(0, 1)$ and represents the fraction of energy of ϕ_k that lies within the frequency band $[-\frac{W}{2}, \frac{W}{2}]$. It is a well-known rule of thumb (mathematically quantified in [7]) that the space of time- and approximately band-limited functions is "approximately WT-dimensional", which can be restated to assert that (approximately) the first $\lfloor WT \rfloor$ eigenvalues $\lambda_1, \ldots, \lambda_{\lfloor WT \rfloor}$ are ≈ 1, and the rest are ≈ 0, with these approximations becoming exact in the limit at $WT \to \infty$.

In addition to the requirement that s be approximately band-limited, a choice of additional constraint(s), such as those described above, will be imposed, which we denote generally as $\Phi_m(s) > c_m$, where m serves as an index to allow multiple independent constraints, if desired. For example, $\Phi_1(s)$ may be $\mathsf{SINR}_{\text{opt}}(s)$ and c_1 an appropriate minimum-acceptable value for radar use. For the applications considered here, the Φ_m are C^∞ functions of s; Most real-world application constraints can be expected to also have nice regularity properties.

Now, let $s \in L^2$ with be expressed as a (complex) linear combination of the prolate spheroidal basis, $s = \sum_{k=0}^{\infty} s^k \phi_k$ with $s^k \in \mathbb{C}$. The total energy of s is $\|s\|_2^2$ and the energy of s within the frequency interval $[-\frac{W}{2}, \frac{W}{2}]$ is $\|Ps\|_2^2 = \langle s, P^* P s \rangle = \sum_k |s^k|^2 \lambda_k$. Therefore, requiring that s leak less than $\delta_0 \ll \|s\|_2^2$ energy out of band is equivalent to the following constraint on the s^k:

$$\sum_{k=0}^{\infty} |s^k|^2 (1 - \lambda_k) \delta_0^{-1} < 1$$

This can be viewed geometrically as requiring that $(s^k) \in \ell^2$ lie inside an infinite-dimensional ellipsoid whose principal axes are the ϕ_k directions, with kth intercept given by $\left(\frac{\delta_0}{1-\lambda_k}\right)^{\frac{1}{2}}$. For any s satisfying (5.2) and any $K \geq 1$, let $\Pi_K s \equiv \sum_{k=1}^{K} s^k \phi_k$ be the orthogonal projection of s onto $\mathcal{S}_K \equiv \text{Span}\{\phi_1, \dots, \phi_K\}$. Since the λ_k are decreasing in k, we have

$$\|s - \Pi_K s\|_2^2 = \sum_{k=K+1}^{\infty} |s^k|^2 \leq \frac{\delta_0}{1 - \lambda_K}$$

In particular, if we choose $K \gtrsim WT$, so that $\lambda_K \ll 1$, then all s satisfying (5.2) are within an L^2 distance of $\lesssim \sqrt{\delta_0}$ of the K-dimensional complex ellipsoid obtained from

(5.2) by setting $s^k = 0$ for $k > K$. Clearly, the information capacity of the radar can only decrease if we restrict transmit waveforms s to lie in this K-dimensional subspace. One can use (5.2), plus the continuity of the constraints Φ_m, to derive an upper bound on the original capacity in terms of a reduced K-dimensional capacity with slightly relaxed values for c_m.

However, in applications to band-limited transmitters and receivers, there is a cleaner approach; Operations of a band-limited system should not rely on being able to detect small signal perturbations outside their nominal spectrum band, which, it must be assumed, could contain significant interference. Since ϕ_k for $k \gtrsim WT$ are overwhelmingly concentrated outside the allowed band, they should not be used for signal construction. An appropriate value K may be determined by examination of the (numerically computed) λ_k associated to the product WT.

Now we can state the capacity problem of interest precisely. Let $K \gtrsim WT$ and $\Phi_m(s)$, c_m be specified for $m \geq 1$. Define $\Phi_0(s) \equiv \|P \circ \Pi_K s\|_2^2 = \sum_{k=0}^{K} |s^k|^2 \lambda_k$, and let $c_0 = 1 - \delta_0$ be the minimum fraction of transmit power required to fall in the specified frequency band. Define the following sets

$$\Omega = \{s \in \mathcal{S}_K : \|s\|_2 = 1, \Phi_m(s) > c_m \text{ for all } m \geq 0\}$$
$$\mathcal{X} = \mathbb{R}^+ \times \Omega = \{s \in \mathcal{S}_K : \Phi_m(s) > c_m \text{ for all } m \geq 0\}$$

Ω is an open subset of $S^{2K-1} = \{\mathcal{S}_K : \|s\|_2 = 1\}$, the unit sphere in $\mathbb{C}^K \approx \mathbb{R}^{2K}$. As such, it is a smooth submanifold of $\mathbb{R}^N$ ($N = n = 2K$), of (real) dimension $2K - 1$. By Theorem 2.1.3 it is in fact a submanifold-with-boundary for Lebesgue-almost-every choice of realizable values of c_m.

We define a communication channel $\mathcal{X} \to \mathcal{Y} = \mathbb{R}^N$ with AWGN of variance $\Sigma =$

$\varepsilon^2 I_N$. We wish to estimate the capacity, subject to some average power constraint on $\mathbb{E}|X|^2 \leq n\mathsf{P}_a$. Our main theorem shows that for $\mathsf{P}_a \gg \varepsilon^2$, this channel capacity is

$$\mathsf{Cap}(\varepsilon) \approx K \log\left(1 + \frac{\mathsf{P}_a}{\varepsilon^2}\right) + \log \frac{V^{2K-1}(\Omega)}{V^{2K-1}(S^{2K-1})}$$

Thus, the constant term $\log \frac{V^{2K-1}(\Omega)}{V^{2K-1}(S^{2K-1})}$ may be considered a zeroth-order correction of the standard K-dimensional Gaussian channel capacity, necessary to account for the constraints Φ_m.

5.3 Numeric Application

In this section we choose realistic radar parameters and numerically compute the ratio $\frac{V^{2K-1}(\Omega)}{V^{2K-1}(S^{2K-1})}$ in order to estimate how the channel capacity varies with choice of waveform constraint parameters. Our numerical methodology is detailed below.

For each chosen time-bandwidth product WT, we take $K = \lceil WT \rceil$. In MATLAB we store the first K discrete prolate spheroidal sequences associated to WT, sampled in time at a frequency that is large compared to W. This is an orthonormal basis spanning the K-dimensional subspace from which all candidate radar transmit waveforms are drawn. We take the constraint $c_0 = 0.92$, so that any candidate waveform having more than 8% of its energy leak out of band is rejected. This value was chosen to agree with the energy leakage of a standard "chirp" waveform of similar time-bandwidth product. We use either $\Phi_1(s) = \mathsf{SIR}_\alpha(s) \geq \beta$ or $\Phi_1(s) = \mathsf{S}_\beta(s) \geq \alpha$, as indicated below, for appropriate values of α, β. The values for minimum signal α are indicated below in absolute terms, with $\alpha \lesssim 1$ desirable (α by definition cannot exceed 1). Values for β are indicated below relative to the equivalent SIR achieved by a standard chirp waveform of the same time-bandwidth product, in order to directly compare performance with a well-known and commonly-used radar waveform.

We compute $\frac{V^{2K-1}(\Omega)}{V^{2K-1}(S^{2K-1})}$ by Monte Carlo simulation. We draw 10,000 random vectors sampled from the distribution $\mathcal{N}(0, I_{2K})$, and normalize the results to obtain uniform-random sampling of S^{2K-1}. The constraints Φ_0 and Φ_1 are evaluated on each sample, and $\frac{V^{2K-1}(\Omega)}{V^{2K-1}(S^{2K-1})}$ is estimated by the fraction of samples which satisfy the constraints.

The functions $\mathsf{SIR}_\alpha(s)$ and $\mathsf{S}_\beta(s)$ must be evaluated numerically. They may be computed efficiently, as we now explain. For $\lambda \in [0,1]$ we define the family of unit-normed filters

$$w_\lambda := \frac{\mathcal{R}_\lambda^{-1} s}{\left\|\mathcal{R}_\lambda^{-1} s\right\|_2}$$

Using the calculus of variations with Lagrange multipliers, it can be shown (e.g., [4]) that the $\{w_\lambda\}_{\lambda\in[0,1]}$ form a family with the following useful properties: For any given λ, w_λ maximizes $\mathsf{S}_s(w)$ among all non-zero w that satisfy $\mathsf{SIR}_s(w) \geq \mathsf{SIR}_s(w_\lambda)$. Conversely, w_λ maximizes $\mathsf{SIR}_s(w)$ among all non-zero w that satisfy $\mathsf{S}_s(w) \geq \mathsf{S}_s(w_\lambda)$. Finally, $\mathsf{S}_s(w_\lambda)$ increases monotonically with λ, and $\mathsf{SIR}_s(w_\lambda)$ decreases monotonically with λ.

Thus $\mathsf{SIR}_\alpha(s)$ may be efficiently computed as follows: compute w_λ and $\mathsf{S}_s(w_\lambda)$ for some initial choice of λ. If the choice of α is achievable, finding the λ^* satisfying $\mathsf{S}_s(w_{\lambda^*}) = \alpha$ is a root-finding problem for a monotonic function in a single variable– a very straightforward task. An $\alpha \in [0,1]$ fails to be achievable only if $\mathsf{S}_s(w_0) > \alpha$, in which case $\mathsf{SIR}_\alpha(s) = \mathsf{SIR}_s(w_0)$. A similar procedure allows computation of $\mathsf{S}_\beta(s)$ for appropriate values of β. Furthermore, the matrix inversion required to compute w_λ may be done once for all λ via a single diagonalization of $\mathcal{R}_0$.

5.3.1 $\mathsf{SIR}_\alpha(s)$ Constraint

Here we take $WT = 10$ and $\Phi_1(s) = \mathsf{SIR}_\alpha(s)$ for several reasonable choices of α. For each α we plot the sample distribution of SIR_α computed from the Monte Carlo simulation. The signal-to-interference ratio is expressed in decibels relative to β_{chirp}, the SIR of a chirp waveform under matched filter, for $WT = 10$. With this information we compute the asymptotic loss of channel capacity $\log \frac{V^{2K-1}(\Omega)}{V^{2K-1}(S^{2K-1})}$ (relative to the equivalent unconstrained Gaussian channel) as a function of the minimum $\mathsf{SIR}_\alpha(s)$ required, i.e. the choice of c_1 in our constraints. The capacity loss is expressed in bits per transmitted radar pulse. The plot demonstrates the trade-off between channel capacity and the basic characteristics imposed on radar transmit waveforms in order to achieve a desired level of performance.

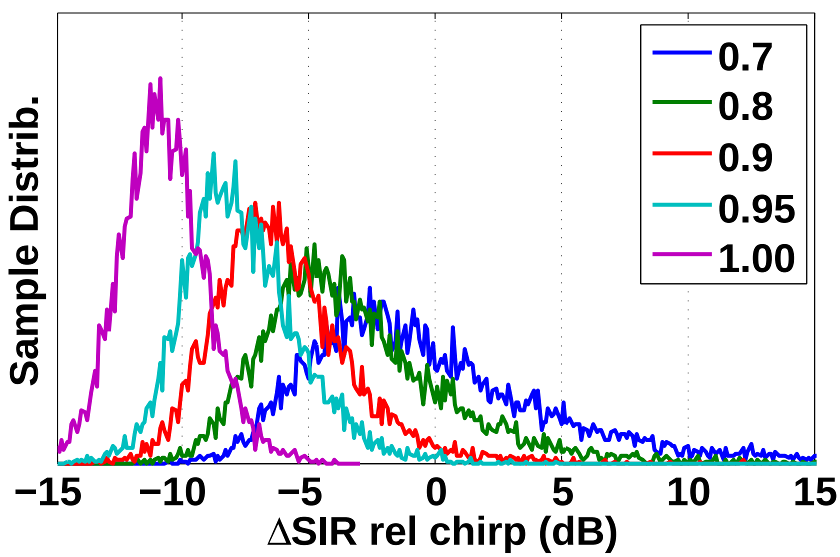

Figure 5.1: Sample distribution of SIR_α for $\alpha \in \{0.7, 0.8, 0.9, 0.95, 1\}$.

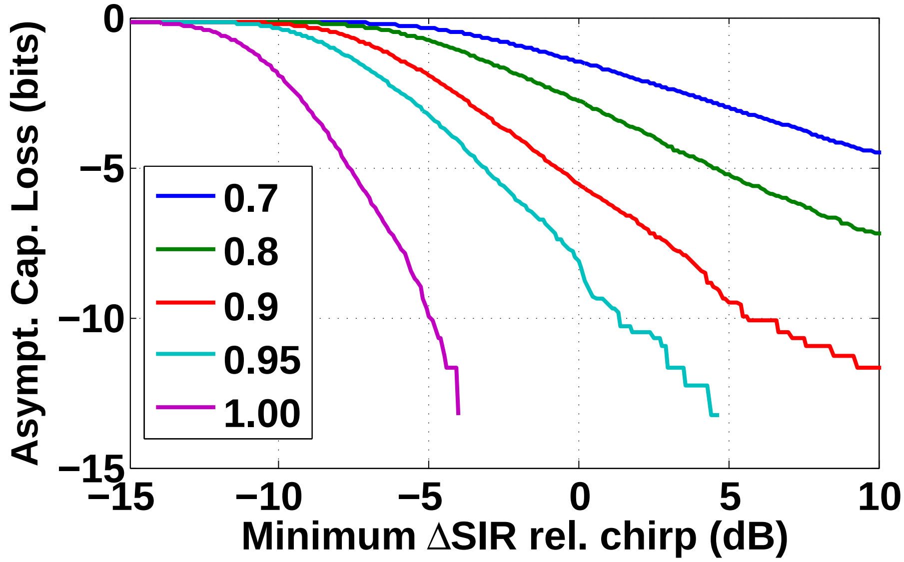

Figure 5.2: Asymptotic capacity loss relative to Gaussian channel for $\alpha \in \{0.7, 0.8, 0.9, 0.95, 1\}$, as a function minimum acceptable waveform SIR.

5.3.2 $\mathsf{S}_\beta(s)$ Constraint

Keeping $WT = 10$, here we take $\Phi_1(s) = \mathsf{S}_\beta(s)$ for a few choices of β, taking the values $\{-5, -3, 0, 3, 5\}$ in dB relative to the SIR of the chirp with matched filter, β_{chirp}. The resulting sample distributions are plotted. The asymptotic capacity loss is plotted as a function of the minimum signal c_1 (expressed in dB, with negative values indicating loss relative to the matched filter). At the left end of the plot we see that nearly all candidate waveforms have a filter that achieves any of the desired SIR. However, moving towards the right end of the chart we see that, if we require the filter to simultaneously maintain a very high signal, the fraction of acceptable waveforms (hence asymptotic capacity) drops precipitously.

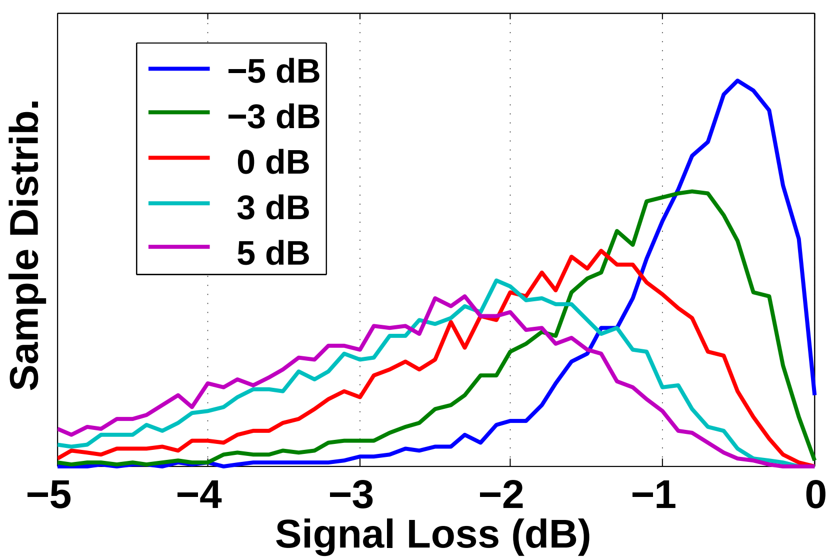

Figure 5.3: Sample distribution of S_β for $10\log_{10}\frac{\beta}{\beta_{\mathsf{chirp}}} \in \{-5, -3, 0, 3, 5\}$.

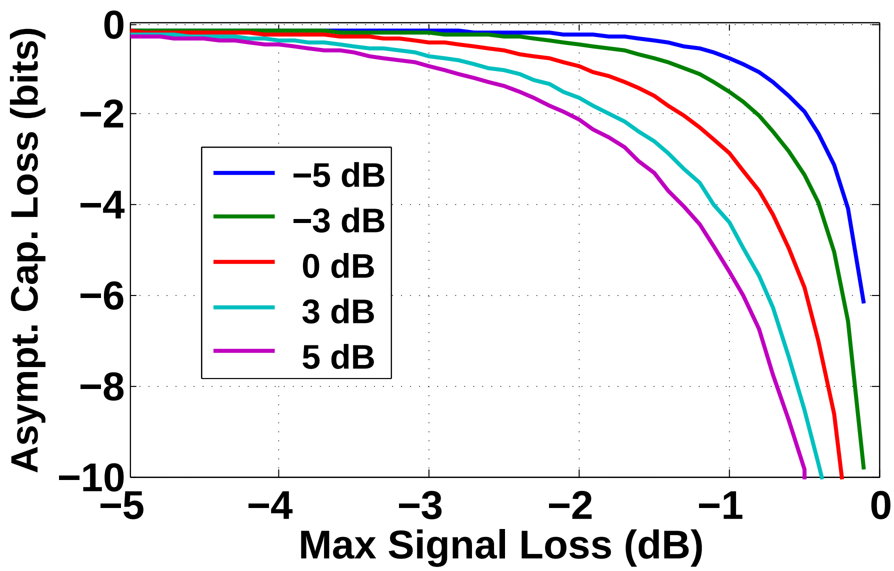

Figure 5.4: Asymptotic capacity loss relative to Gaussian channel for $10\log_{10}\frac{\beta}{\beta_{\text{chirp}}} \in \{-5, -3, 0, 3, 5\}$, as a function minimum acceptable waveform S.

5.3.3 $\mathsf{S}_{\beta_{\text{CHIRP}}}$ Constraint with varying time-bandwidth product

Finally, we take several choices of time-bandwidth product $WT \in \{10, 20, 30, 50, 100\}$ and take $\Phi_1 = \mathsf{S}_{\beta_{\text{chirp}}}$, where β_{chirp} is the SIR of the chirp waveform and matched filter corresponding to the chosen WT. Our plot of asymptotic capacity loss is normalized by the nominal degrees of freedom WT for the sake of comparison. With this normalization, the asymptotic capacity loss curve appears to be largely independent of time-bandwidth product.

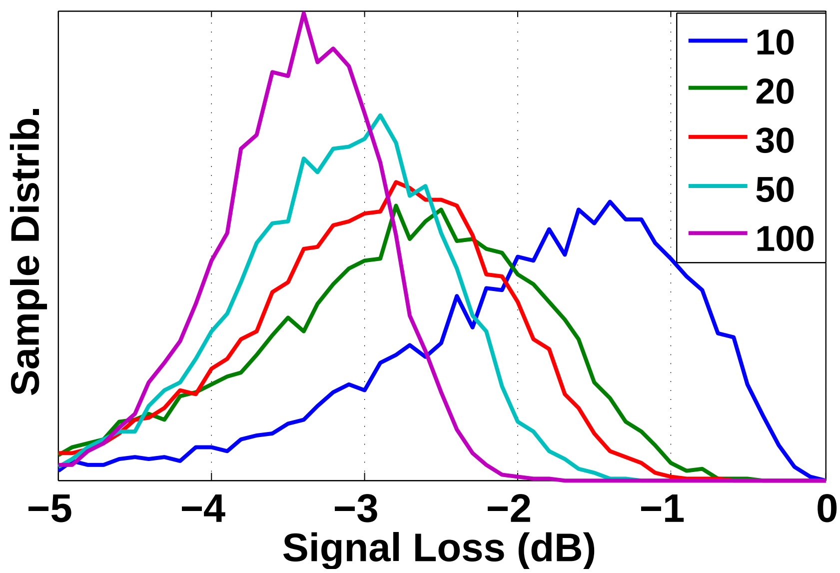

Figure 5.5: Sample distribution of $\mathsf{S}_{\beta_{\text{chirp}}}$ for $WT \in \{10, 20, 30, 50, 100\}$.

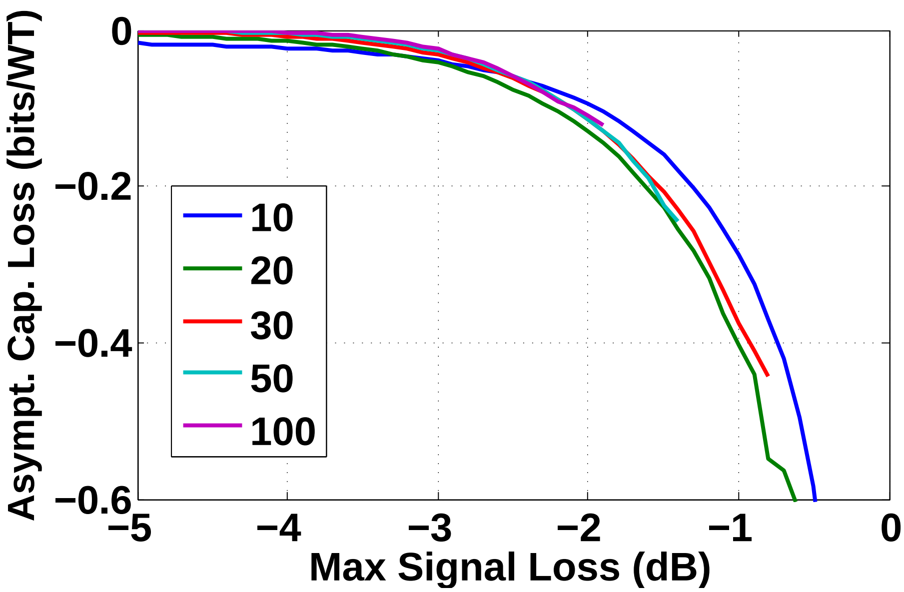

Figure 5.6: Asymptotic capacity loss per degree of freedom, relative to Gaussian channel for $WT \in \{10, 20, 30, 50, 100\}$.

6

Summary and Next Steps

Our work in this paper approaches two fundamental quantities of Information Theory – entropy and capacity – from the mathematical perspectives of Real Analysis and Differential Geometry, to develop novel estimation theorems with quantifiable error bounds. We have shown the power of the general theory we have developed by the application to the radar waveform capacity problem; The constraint equations imposed by this problem seem extremely difficult even to write down and manipulate in closed form, let alone to analyze for the purposes of computing an approximate channel capacity. The combination of our general asymptotic capacity analysis and a straightforward and efficient numerical routine allows us to elegantly sidestep this difficulty entirely, at least in the high-SNR regime, to quantify the trade-off between effective radar operation and communication capacity.

We believe there are a number of other alphabet-constrained capacity problems which are not amenable to a traditional capacity analysis which may be approached using the techniques developed here. For example, a channel of broad interest is the complex AWGN channel with both an average and a peak power constraint imposed (see, for example, [11]). The peak power constraint may be considered in our context to be an alphabet constrained to the closed ball $\bar{B}^2$. Our current capacity results (Theorem 4.2.2 and Theorem 4.3.1) do not directly apply to this case, which contains elements of both the compact case and the average-power-constrained case. We believe that, with follow on work, it may be possible to apply our general asymptotic mutual information approximation theorem (Theorem 4.2.1) to analyze this case as well.

We have been careful in this paper to develop estimates with quantifiable, computable error bounds to the greatest degree possible. While our final results are stated in terms of unspecified constants, it is possible, with a significant amount of work, to compute an *explicit* error bound for our capacity estimate, thus converting an asymptotic high-SNR result into an explicit range inside which the *exact* channel capacity is thus proven to lie. The bound will depend on the quantities specified throughout the computation– notably, the geometric bounding constant $c_{\mathcal{X}}$, which may be difficult to compute in many cases, including the radar waveform capacity problem. New numerical techniques would be required here. However, it is not hard to compute $c_{\mathcal{X}}$ for more explicitly described alphabet manifolds such as the closed ball $\bar{B}^2$. In either case, future work towards the goal of explicit error bounds would only be of use if the resulting capacity error bounds are typically small enough to give a non-trivial range of possible capacities. This could require a careful accounting of error terms and exacting work attempting to estimate the induced errors as accurately as possible. It is not a trivial exercise, but the ability to compute meaningful, exact capacity ranges for a variety of SNR in our general setting would be an exciting development.

Another direction for refinement of our present results is the computation of higher order terms in ε for our asymptotic approximation. A heuristic argument suggests that we might view the input manifold $\mathcal{X}$ as approximated by small pieces of n-spheres (and possibly other simplified spaces), whose radii are determined by the curvatures of $\mathcal{X}$ in that area. This suggests that a higher order approximate capacity achieving input distribution will vary with the local curvature of $\mathcal{X}$. We believe that an expression for one or more additional higher order terms in ε may be computable, at least for sufficiently tractable geometries, but not without significant additional research.

Finally, we limited our radar waveform capacity investigation to a single pulse radar model for computational simplicity. In principle our approach can be extended to other forms of radar. In particular, a *pulse-Doppler radar*, which emits a coherent set of M pulses in order to process target Doppler from pulse-to-pulse phase shifts, is of practical interest. Our methodology may be extended to this case with little additional theoretical analysis. However, as the M pulses must be processed coherently for Doppler information, the natural application of our framework would be to consider all M pulses together as a single code letter in a larger dimensional ambient space. Appropriate additional alphabet constraints would need to be considered to ensure effective Doppler processing. Finally, while the numerical approach used in this paper to compute capacity extends easily to the pulse-Doppler radar scenario, the dimensionality increases by a factor of M. In fielded pulse-Doppler radars M can range anywhere from < 10 to > 1000, depending on the required Doppler resolution. This will have a significant impact on time required for the computation, although it is certainly a tractable computation for modern institutional computational resources.

References

[1] Chiriyath, A. R., Paul, B., Jacyna, G. M., & Bliss, D. W. (2016). Inner bounds on performance of radar and communications co-existence. *IEEE Transactions on Signal Processing*, 64(2), 464–474.

[2] do Carmo, M. P. (1976). *Differential geometry of curves and surfaces*. Prentice-Hall.

[3] do Carmo, M. P. & Flaherty, F. J. (1992). *Riemannian geometry*. Birkhäuser.

[4] Gilbert, E. N. & Morgan, S. P. (1955). Optimum design of directive antenna arrays subject to random variations. *Bell System Technical Journal*, 34(3), 637–663.

[5] Gray, A. (2004). *Tubes*. Birkhäuser Basel, 2 edition.

[6] Jamil, M., Zepernick, H.-J., & Pettersson, M. I. (2008). On integrated radar and communication systems using oppermann sequences. *MILCOM 2008 - 2008 IEEE Military Communications Conference*.

[7] Landau, H. J. & Pollak, H. O. (1962). Prolate spheroidal wave functions, fourier analysis and uncertainty-iii: The dimension of the space of essentially time- and band-limited signals. *Bell System Technical Journal*, 41(4), 1295–1336.

[8] Nijsure, Y., Chen, Y., Yuen, C., & Chew, Y. H. (2011). Location-aware spectrum and power allocation in joint cognitive communication-radar networks. *Proceedings of the 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications*.

[9] O'Neill, B. (1983). *Semi-Riemannian geometry: with applications to relativity*. Academic Press.

[10] Sard, A. (1942). The measure of the critical values of differentiable maps. *Bull. Amer. Math. Soc. Bulletin of the American Mathematical Society*, 48(12), 883–891.

[11] Shamai, S. & Bar-David, I. (1995). The capacity of average and peak-power-limited quadrature gaussian channels. *IEEE Transactions on Information Theory*, 41(4), 1060–1071.

[12] Shannon, C. E. (1959). Probability of error for optimal codes in a gaussian channel. *Bell System Technical Journal*, 38(3), 611–656.

[13] Weyl, H. (1939). On the volume of tubes. *American Journal of Mathematics*, 61(2), 461.

[14] Zheng, L. & Tse, D. N. C. (2002). Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Transactions on Information Theory*, 48(2), 359–383.